Volume 5 SEPTEMBER, 1934 Number 9

PHYSICS

A Journal of General and Applied Physics

Nonmetallic Conducting Films*

S. BLOOMENTHAL, *U. S. War Department, Frankford Arsenal, Philadelphia, Pa.*†

(Received May 7, 1934)

Immutable conducting films, which adhere to any clean solid have been formed from synthetic resin solutions holding finely ground carbon in suspension. Values of surface resistivity for uniformly thin films were secured in the range 15 to 10^6 ohms referred to unit area of surface. When the thickness is held constant, the resistivity depends mainly upon the proportion of binder present and the kind of carbon used. The experimental error in repeating values of resistance with a given suspension is about ±20 percent. The dependence of resistance upon resin content of the film appears to be in accord with the views of Frenkel concerning gap contacts and in definite disagreement with the classical Maxwell theory of composite conductors. The behavior of various types of carbon, when used in the films, can be predicted qualitatively from electrical measurements made upon powdered samples subjected to high pressure. Absorption of moisture by a film leads to a rise in resistance.

INTRODUCTION

ALTHOUGH composite conductors are used as high-value resistance elements in nearly every radio receiving set, there is practically no information in the literature dealing with their properties and formation. The composite conductor may take the form of a film adhering to an insulator or may be used as a disk or cylinder formed under pressure. Conducting carbon is used in nearly all of these resistances.

Many years ago Maxwell[1] considered a composite conductor composed of spheres distributed within a continuous conductor and worked out an equation relating the resultant resistivity with the resistivities and volumes occupied by the components. Later Rayleigh[2] attacked the problem from a slightly different angle and considered the case of a rectangular array of uniform insulating spheres in a conducting matrix and indicated what correction must be applied to the simple Maxwell formula when the volume occupied by the spheres is large. Rayleigh's deductions have been verified experimentally by Wold[3] for the case of mercury containing gas bubbles.

The patent literature dealing with the processing of carbon type radio resistors is fairly extensive and an attempt to do justice to this subject would lead us far from the object of the present paper. For many years a number of business concerns such as the Allen Bradley Company, the Stackpole Carbon Company, the Erie Resistor Company, the International Resistor Company, the Chicago Telephone Supply Company, the Continental Carbon Company and others have been manufacturing and marketing radio resistors. For the film type fixed or variable resistor a suspension of finely divided carbon in cellulose acetate or nitrate, shellac, natural resin, drying oil or synthetic resin is used. In making the best type of molded fixed resistor, a carbon

* Paper presented at the Cincinnati Meeting of The Am. Phys. Soc., Dec. 1, 1933. Phys. Rev. **45**, 122(A) (1934).

† This work was done in the laboratories of the Radio Corporation Of America at Camden, New Jersey.

[1] J. C. Maxwell, *Electricity and Magnetism I*, p. 425 (2nd edition).

[2] Rayleigh, *Scientific Papers* **IV**, 19 (1892).

[3] P. I. Wold, Phys. Rev. **43**, 383 (1933).

suspension in a synthetic resin solution with silica, mica dust, iron oxide or asbestos fibers as the inactive filler is employed. After removal of the solvent by evaporation, the pulverized material is cold-molded in tabletting machines and baked to effect polymerization of the resin binder. Metal electrodes are sprayed on the cylindrical resistors.

The nonmetallic conducting films with which this report deals were formed from suspensions of finely divided conducting carbon in phenol formaldehyde resin solutions. Even though the carbon particles are completely separated, one from the other, by resin layers after the films are dried and baked, conduction of electricity takes place through the medium. According to the wave-mechanical theory of Frenkel[4] concerning metallic gap contacts, the resistivity should rise very rapidly as the spacing between contacts increases. With carbon crystals having a positive temperature coefficient of resistivity[5] one ought to expect a similar behavior. Carbon is soft and of low density in its common form and can be ground in a liquid to give suspensions which do not settle out rapidly. When heated, carbon does not form poorly conducting solid oxide layers. In certain practical applications the self-lubricating qualities due to the presence of graphite are essential. One has a wide choice of carbon to use in controlling film resistivity characteristics.

Insulating phenolic resins of the Bakelite type[6] are thermosetting at moderately high temperatures (100 to 200°C) and give some assurance of resistance stability in the composite conductor. It is well known that certain varieties of synthetic resins have only a slight tendency to absorb moisture when in the cured condition,[7] shown by measurements on the decrease of resistance when exposed to humid conditions. Commercially available forms are readily soluble in a number of organic liquids.

Nonmetallic conducting films are chemically inert and the writer has used them as a coating on the aluminum cathode of an electrolytic condenser in order to prevent the drop in capacity with use due to metallic oxide film formation on the cathode in the conventional design.

Electrical Behavior of Powdered Carbon Subjected to Pressure

The resistivity characteristics of carbon powders can be determined from electrical measurements made upon samples placed in turn within an insulating cup and subjected to pressure. The piston may serve as one of the electrodes. Such measurements have been made in the past chiefly by E. Ryschkewitsch[8] and Z. Specht.[9] They found that graphite powders, both natural and artificial, obeyed the law:

$$\alpha = a/p + b, \qquad (1)$$

where α is the resistivity, a, a constant of value 0.3 to 36.5, b, a constant of value 0.0075 to 0.14 and p the pressure in kg/cm² lying between 15 and 234 in value. The low values of the constants apply to the pure varieties of graphite and the high values to the impure varieties. These experiments appear to be in agreement with the theoretical and experimental work of Pederson[10] on carbon.

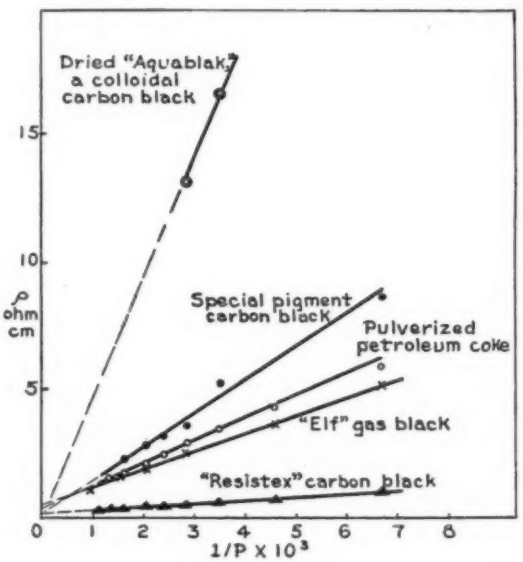

Fig. 1. Variation of resistivity of amorphous carbon powders with reciprocal of the applied pressure in kg/cm².

[4] J. Frenkel, Phys. Rev. **36**, 1604 (1930).
[5] D. E. Roberts, Ann. d. Physik **40**, 453 (1913); Phil. Mag. **26**, 158 (1913).
[6] L. V. Redman, Int. Crit. Tab. II, p. 299.
[7] H. L. Curtiss, Bull. Bur. Stand. **11**, 359 (1915).
[8] E. Ryschkewitsch, Zeits. f. Elektrochem. **28**, 289 (1922); **29**, 274 (1923).
[9] Z. Specht, Zeits. f. Elektrochem. **38**, 920 (1932).
[10] P. O. Pederson, The Electrician, **76**, 625 (1916).

In order to get a qualitative anticipation of the comportment of specimens of powdered carbon before they were used in conducting films, the writer constructed a measuring apparatus consisting of a steel jacketed Bakelite cup having a brass electrode in its base. A copper piston served also as the upper electrical contact. A Leeds and Northrup type K potentiometer measured the potential drop across the cup as the pressure was varied, while a small constant value of current flowed through the carbon in the cup from a battery. Pressure varying in steps of about 60 kg/sq. cm was applied to the sample by means of an air press. As the pressure increased the resistance and potential difference measured decreased. Fig. 1 illustrates the variation of resistivity with pressure for a number of samples of amorphous carbon, which have a high specific resistivity in general. Lampblacks conduct fairly well, however, and stand between the best conducting gas carbon and the most poorly conducting graphite. Fig. 2 shows the curves for graphite powders. The samples were all finely pulverized and contained few particles coarse enough to be retained upon a 150-mesh screen. The commercial source of some of these carbon samples will be given in the section of this paper dealing with their use in conducting films. The resistivities of all the samples of graphite fell below one ohm cm, with considerable difference in behavior between the artificial varieties and natural forms, of which Mexican graphite is one representative.

In a recent paper Brunner and Hammerschmid[11] give the results of measurements of resistivity made on six different kinds of graphite subjected to pressures of 31 to 300 atmospheres and find that the resistance varies inversely as the square root of the pressure. These results are at variance with most of the published data.

Metering Suspensions by the Drop Weight Method

The conducting films are formed by the evaporation of solvent from a synthetic resin solution acting as vehicle for the suspended carbon particles. Comparison between the drop

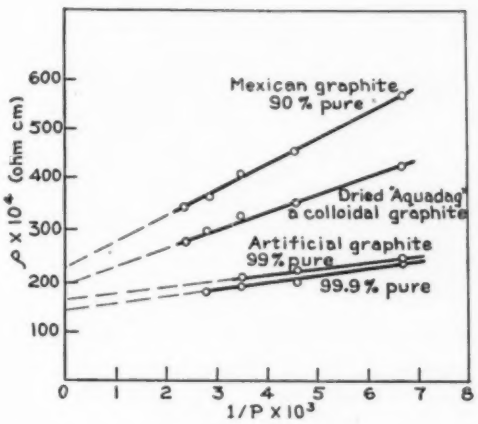

FIG. 2. Variation of resistivity of graphite powders with reciprocal of the applied pressure in kg/cm².

weight of the pure solvent and a fresh suspension containing 20 percent by weight in solids (resin +carbon) when drops of the two were formed from the same orifice, gave a difference of less than two percent. This indicated that the drop weight is governed principally by the surface tension of the liquid used as the solvent for the resin.

The drop weight M is a function of the density of the liquid O, the surface tension T, the radius of the dropping tube a and the acceleration due to gravity g and has been determined approximately by Rayleigh[12] from dimensional analysis as

$$M=(Ta/g)F(T/gOa^2),$$

where F denotes an arbitrary function. For the case of water drops Rayleigh found F to be 3.8, a value near that given much later by Harkins and Young.[13] The writer constructed a brass dropper with external diameter of 0.0932 inch and with wall thickness of 0.0156 inch and used it in the present work. With water the drop weight was 0.032 gram, a value similar to 0.035 gram calculated from Rayleigh's equation. If the liquid used is of low volatility and viscosity, one can measure small quantities fairly accurately with a dropper. This is the basis for the present method of producing films of fairly definite resistivity characteristics.

[11] J. Brunner and H. Hammerschmid, Zeits. f. Elektrochem. **40**, 60 (1934).

[12] Rayleigh, reference 2, p. 415.

[13] T. F. Young and W. D. Harkins, Int. Crit. Tab. IV, p. 435.

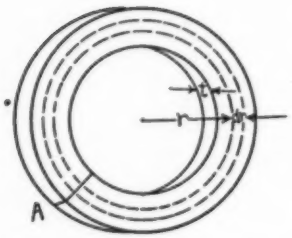

FIG. 3. Geometrical form of the conducting film.

THE DIELECTRIC BASE UPON WHICH THE FILMS ARE DEPOSITED

For convenience in application of the suspensions, an annular depression in the surface of a molded disk was employed to retain the film forming liquid in most of the experiments. When the solvent evaporates, the resistance layer is of the form shown in Fig. 3. Imagine a narrow cut through the element, and electrodes attatched to the walls of the cut at A. dr is the width of an elementary conductor of the ring of thickness t. If ρ is the volume resistivity of the composite conductor and R the resistance,

$$R=\left(\int_{r_1}^{r_2}\frac{t\,dr}{2\pi r\rho}\right)^{-1}=\frac{2\pi\rho}{t\log r_2/r_1}. \qquad (2)$$

Eq. (2) should give the same result as that for a linear conductor when r_1, the inner radius, and r_2, the outer radius, of the annulus are both large and nearly equal. Thus with $r_2=100$ cm and $r_1=99$ cm, the value of R in terms of ρ/t calculated from the two formulas differ by only 5 percent. The derivation of (2) follows the method of Thomson[14] who found the resistance from the inner to the outer wall of a cylinder. The values of r_1 and r_2 were 1.15 and 1.85 cm, respectively. t was of the order 0.001 inch.

The disk-shaped bases were molded in a positive type mold from a plastic compound prepared on hot mixing rolls maintained at about 225°F. The composition by weight was 35 percent phenol-formaldehyde resin, G. E. Co. brand, 64.8 percent powdered glass sieved through a 200 mesh screen and in addition, 0.2 percent montan wax to prevent sticking of the molded piece to the surfaces of the mold. The compound was ground to pass a 60-mesh screen and was pressed into disk-shaped preforms, which were then subjected to a pressure of about 600 lb./in.^{2} at 300°F in the mold for six minutes. The mold was water-cooled before ejection. The granular surface required for a slow spreading of the suspensions was secured by etching the molded pieces for 30 seconds in 5 percent NaOH held at the boiling point. The bases were then washed in 2 percent HCl and in boiling water. The cure was then completed by baking at 130°C for 24 hours. This precaution is necessary because the generation of gas within the bases at 170° necessary for formation of conducting films causes distortion and blistering if the cure is incomplete. Joints between films can be made easily upon the etched surface because excessive flow of liquid by capillarity to the partially dry film made first is prevented.

[14] J. J. Thomson, *Elements of Electricity and Magnetism*, I, p. 242 (1921).

EFFECT OF SEDIMENTATION UPON FILM RESISTANCE

During preliminary experiments an attempt was made to determine the accuracy with which resistance values of films could be repeated. One hundred fifty-two films were formed by the drop method by a female assistant, selected from several hundred workers in a condenser winding factory because of her reputation for exceptional accuracy. The resistance sections formed upon 100° of the annulus of Fig. 3 were numbered consecutively from 777 to 929. The resistance value of each section is represented by a dot in

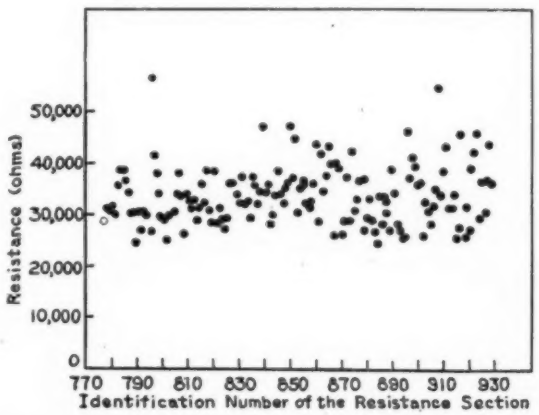

FIG. 4. Spreading of film resistance values due to settling of the suspensions. Sections Nos. 777 to 806 made during one day, a Thursday, 807 to 832 during the next, 833 to 875 on the Monday following, and 875 to 929 during the next day.

Fig. 4. Segments Nos. 777 to 806 were made during one day, a Thursday, segments No. 807 to 832 on the next day; Nos. 833 to 875 were made on the Monday following and segments Nos. 875 to 929 on the next day. The spreading of the resistance values due to settling of the suspensions is evident. When first formed, the carbon suspensions are fairly stable, but after a few days sedimentation sets in and homogeneity cannot be restored by a mere shaking by hand of the bottle containing the suspension. The low values of resistance secured after sedimentation is pronounced may be due to quickness on the part of the operator in withdrawing the fluid from the bottle with the dropper. The high values of resistance, shown especially in the right-hand side of the figure are due both to lack of solid matter in the fluid or unbalance between the amounts of the two kinds of carbon present due to sedimentation.

The suspension used in this test had the same composition as that used to give the data of line 1, Series 3 of Table I of the next section. The same general technique was used also.

Dependence of Film Resistivity Upon Composition

In order to secure the lowest possible film resistivity, one uses a crystalline graphite with just enough resin to give strength and adhesion,

Table I. *Variation of film resistivity with composition. Series 1 has phenolic resin 65 percent by weight, graphite 35 percent (two types used). Series 2 shows the effect of replacing part of the graphite present in 1 by a component of higher resistivity than graphite. Phenolic resin 65 percent, graphite 17.5 percent and third component 17.5 percent. Series 3 shows the variation in resistance produced by substitution of one type of graphite by another. Phenolic resin 65 percent, gas carbon 17.5 percent and graphite 17.5 percent. Series 4 gives the increase of resistance with resin content with constant ratio of 1 : 1 by weight maintained between the components crystalline e.f. graphite and gas carbon. Series 5 gives similar data with the graphite of the preceding series substituted by an amorphous natural variety. Series 6 gives data for films containing crystalline e.f. graphite and gas carbon in the ratio 1 : 2 by weight. In Series 7 amorphous natural graphite and resin are combined to give high value resistance films. The small change in the value of resin density due to polymerization is taken into account in the calculation of ρ.*

Series	*R* Measured (ohms)	*S* (ohms)	ρ (ohm cm)	Drop wt. (g)	Wt. of film (g)	
						Type of Graphite Used
1. *a.*	200	16.3	0.088	0.0129	0.0312	Crystalline natural
b.	1330	145	0.572	0.0122	0.0266	Crystalline electric furnace
						Third Component
2. *a.*	4110	446	1.98	0.0126	0.0285	Lampblack
b.	112,000	12,200	52.1	0.0119	0.0286	Silica
						Type of Graphite Used
3.	37,500	4070	12.6	0.0126	0.0202	Crystalline e.f. (Acheson No. 1340, 99 percent C)
	50,000	5440	16.0	0.0120	0.0192	Amorphous e.f. (Acheson No. 40, 96 percent C)
	60,000	6520	20.2	0.0127	0.0204	Amorphous e.f. (Acheson No. 41, 91 percent C)
	138,000	15,000	44.6	0.0124	0.0198	Amorphous natural (Dixon's No. 0708, 90 percent C)
						Percent Phenolic Resin
4.	37,500	4070	12.6	0.0126	0.0202	65.0
	43,400	4710	12.0	0.0132	0.0163	70.0
	133,000	14,500	37.6	0.0133	0.0164	72.5
	484,000	52,600	127.0	0.0122	0.0150	75.0
5.	104,000	11,300	29.7	0.0109	0.0175	60.0
	138,000	15,000	44.6	0.0124	0.0198	65.0
	260,000	28,300	89.6	0.0128	0.0205	67.0
	900,000	97,800	245.0	0.0120	0.0161	69.0
6.	47,500	5160	12.6	0.0125	0.0160	70.0
	1,900,000	205,000	504	"	"	72.5
	9,000,000	978,000	2420	"	"	75.0
7.	5612	610	1.02	0.0120	0.0116	57.2
	73,000	7940	14.2	"	"	72.5
	1,400,000	152,000	282.0	"	"	80.0

as in Series 1 of Table I. If part of the graphite is substituted by a less efficient conductor, the film resistance ought to rise and this effect is shown in Series 2. Series 3 shows that a variation in resistance may be produced by substituting one variety of graphite by another when half of the carbon mixture is one kind of gas carbon. The remaining series show that the resistance may be raised by increase in the resin content.

In securing the data of Table I an attempt was made to maintain comparable conditions of film formation. Bases molded and etched in the manner described above were used. The time of preparation of the suspension used for each line of Table I was 24 hours with a speed of 100 R.P.M. for a 1000 cc pebble mill half full of pebbles and charged with 300 g. A carefully levelled tripod supported the bases while eight drops were deposited upon 240 degrees of the annulus of Fig. 3 from the brass dropper. After the main films had dried for a few minutes, low resistance contacting areas consisting of 0.007 g of Series 1 lowest resistance mixture were formed over 30 degrees of the annulus overlapping a small part of each end of the main conducting film. After air drying for fifteen minutes, the films were baked at 110°C for 10 minutes and finally at 170°C for 30 minutes. After cooling, the film was measured with a Leeds and Northrup portable type resistance bridge while metallic contacts were pressed against the low resistance areas. Potentiometer tests showed that the films obeyed Ohm's law for minute values of current.

In Table I ρ, the volume resistivity, is derived from the measured resistance R and the geometry of the film (Eq. 2). Even if one neglects any voids that may be present, he finds that experimental values of thickness compare well with calculated values. The resistance S referred to unit area of the film is ρ/t.[15] The percentage of solids (resin +carbon) in the suspensions was 20 percent except for series 1*a* having 27.4 percent, 1*b* and 2 with 30 percent and 7 with 25 percent. General Plastics Company Durez No. 175 lump phenolic resin was used for all the trials except 1*a* in which Bakelite Corporation No. 3189 resin was used. The solvents consisted of equal parts by weight of acetone and amylacetate for all the experiments except 1*a* in which the ratio was 1 to 1.7, and Series 7 for which alcohol was used. Eimer and Amend "Elf" gas carbon was used in Series 3, 4, 5 and 6, Acheson No. 1340 electric furnace (abb. e.f.) graphite in 1*b*, 4 and 6, Dixon's No. 066 graphite in 1*a* and Dixon's No. 0708 graphite in 7. The silica of 2*b* is Johns Manville Company's "Superfloss" and the lampblack of 2*a*, the variety "Eagle" sold by the L. Martin Company.

In connection with the electrical properties of metallic films produced by evaporation or sputtering, much study has been given to the existence of a critical film thickness below which a new and higher value of resistivity than that of the metal in bulk is present.[16] One might expect to find that when the thickness of a nonmetallic conducting film became comparable with the size of the grains of conducting carbon present that an analogous phenomenon would be apparent. The films used in the experiments described in the paragraphs above were on the average about ten times the particle diameter in thickness. It was found that a doubling of the film thickness halved the resistance, as expected. However when an attempt was made to secure thinner films a rapid rise in resistance was found. With experimental

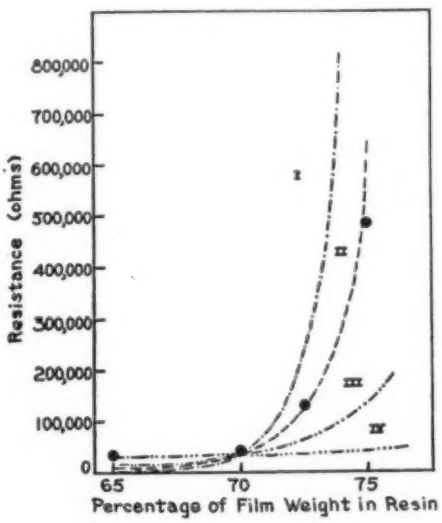

FIG. 5. Dependence of film resistance upon resin content. Dots are experimental values of Series IV. The curves are derived theoretically, I, II and II from Eq. (6) for $\delta = 6A$, 4A and 2A, respectively. IV follows Maxwell's theoretical prediction.

[15] J. C. Maxwell, reference 1, II, p. 287.

[16] L. Hamburger, Ann. d. Physik **10**, 649, 789, and 905 (1931); **11**, 40 (1931); J. Kramer, Ann. d. Physik **19**, 37 (1934).

conditions similar to the previous ones, and with suspensions having the composition phenolic resin 52.5 percent by weight, gas carbon 19.0 and graphite 28.4 but with concentration of solids of 20.2 percent, 10.3, and 6.9 percent, resistance values of 80,000 ohms, 700,000 ohms and 1,600,000 ohms, respectively, were secured. Ethyl alcohol was used as the solvent. This result indicates that very thin nonmetallic conducting films may show a critical thickness below which the volume resistivity is no longer a constant.

Although the experimental data secured were not considered to be of exceptional accuracy, a comparison of the dependence of film resistance upon resin content dealt with in Series 4 of Table I with the predictions of theory was considered worth while. Fig. 4 shows clearly that the accuracy with which resistance values may be duplicated is about ± 20 percent when the suspension is fresh. With a second suspension prepared identically, a considerable shift of the average value may be expected. The results of Series 4 are plotted in Fig. 5 as large dots. The ordinates are resistance values in ohms and the abscissas the percentage of the weight of the film in resin.

The theoretical expression in the form given by Maxwell[17] for a conductor of fixed dimensions with imbedded spheres of infinite specific resistance, when the diameter of the spheres is small compared to the distance between them is

$$R_2/R_1=(1+\tfrac{1}{2}p_1)/(1-p_1), \qquad (3)$$

where R_2 is the resistance with the insulator present, R_1 the value without the insulator and p_1 the fractional volume occupied by the spheres. Eq. (3) is used for curve IV in Fig. 4 with the ratio modified because percentage of insulator by weight rather than by volume is used in the figure. The deviation of curve IV from the experimental data is very great.

Frenkel's theory[18] is based upon the idea that a gap between two adjacent conductors can be considered as a potential hill with vertical slopes at the two surfaces. An electron can leap over the hill according to wave mechanics even though it does not have sufficient kinetic energy to do so according to classical mechanics. His expression relating the conductivity σ_g to the gap distance δ in A is

$$\sigma_g/\sigma=\delta/le^{-2.48}. \qquad (4)$$

σ_g equals δ/R_s with R_s the resistance per unit surface of the contact. σ is the conductor's conductivity. 1 is the mean free path of the electrons, given as 100A. If we imagine the carbon particles in the film to be elementary cubes of edge length d cm arranged in layers with a spacing δ' between faces, we can get some idea of how the resistance should vary with resin content in an ideal case. The dependence of the spacing upon resin content can be secured from the expression

$$O_2[(PW/m)^{\frac{1}{3}}(d+\delta')-\delta']^3=WO_1d^3/m. \qquad (5)$$

In Eq. (5) O_2 is the density of the composite conductor and O_1 that of the carbon cubes, P the fractional weight of the total due to the carbon, W the weight of a cubical sample of the composite conductor, and m the mass of one of the small cubes. Assuming that conduction takes place across the gaps between layers of cubes, one finds

$$R=R'(\delta/\delta')(C'/C), \qquad (6)$$

where R is the film resistance corresponding to an elementary gap width δ, taken at 70 percent resin content and R' the value for a gap width δ'. C and C' are the values of σ_g for gap widths δ and δ'. Eq. (6) is plotted in Fig. 5 for several values of δ.

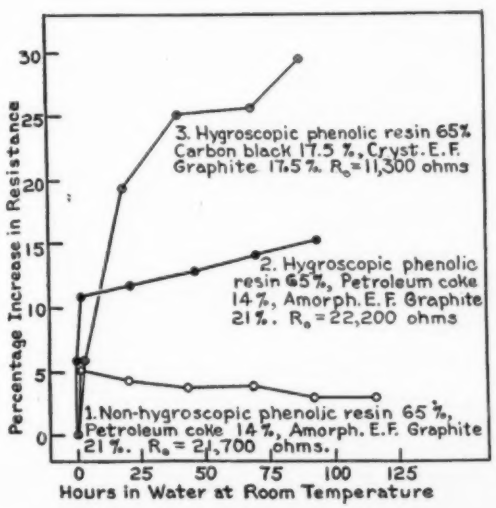
Percentage Increase in Resistance
30
25
20
15
10
5
0
3. Hygroscopic phenolic resin 65% Carbon black 17.5 %, Cryst. E. F. Graphite 17.5 %. R₀=11,300 ohms
2. Hygroscopic phenolic resin 65%, Petroleum coke 14 %, Amorph. E. F. Graphite 21 %. R₀=22,200 ohms
1. Non-hygroscopic phenolic resin 65 %, Petroleum coke 14 %, Amorph. E. F. Graphite 21 %. R₀=21,700 ohms.
0
25
50
75
100
125
Hours in Water at Room Temperature

FIG. 6. Rise in resistance due to moisture absorption for films made from three typical suspensions.

[17] J. C. Maxwell, reference 1, I, p. 435.
[18] J. Frenkel, reference 4.

Increase in Film Resistance due to Absorption of Moisture

Absorption of moisture by a nonmetallic film of the type described invariably leads to a rise in resistance. The magnitude of the rise depends somewhat upon the absolute value of surface resistivity, the type of binder used and the nature of the conducting carbon. In order to study the effect under ideal conditions, the writer used four parallel strips 1×6 cm formed from each suspension upon a glass plate 9×6 cm. A 1×1 cm square on each end of the strips was covered with 0.036 g of the low resistance mixture used in Series 1*a*. The curves of Fig. 6 show how the resistance value varies with time of immersion for films made from three unlike suspensions. For curve 1 a non-hygroscopic insulating phenolic resin varnish, General Plastics Company Durez No. 3127 D is combined with slightly hygroscopic carbon, petroleum coke and Acheson No. 40 e.f. graphite; for 2 a hygroscopic phenolic resin, Durez No. 175 is combined with the same carbon mixture as in 1 and for 3 the same type of binder as for 2 is combined with a carbon mixture of Acheson e.f. No. 38 graphite and hygroscopic Binney & Smith Co. "Resistex" carbon black. The proportions used are shown. R_0 is the initial film resistance. 0.0025 g of material/sq. cm was deposited. The results follow moisture absorption data for resins[19] and various species of carbon powders.[20] R_0 is restored upon drying.

Acknowledgments

To Messrs. W. R. G. Baker, vice-president, G. W. Carpenter, manager of engineering, E. W. Engstrom, research director and Dr. I. Wolff, head of the physics research section, of the RCA Victor Company, the writer is indebted for much helpful advice and encouragement. Thanks are due Dr. A. J. Norton of the General Plastics Company, Dr. L. V. Redman and Mr. C. A. Nash of the Bakelite Corporation, Mr. J. W. Willitts of the Joseph Dixon Crucible Company and Dr. W. B. Wiegand of the Binney and Smith Company for their aid in the selection of raw materials with which to work.

[19] L. V. Redman, reference 6.

[20] D. C. Lindsay, Int. Crit Tab. II, p. 321, coke and carbon black; H. M. Langton, Blacks and Pitches p. 39 (1925) cracked gas carbon; C. L. Mantell, Industrial Carbon, p. 92 (1928) carbon black.

Fluctuation Noise in Vacuum Tubes

G. L. PEARSON, *Bell Telephone Laboratories, New York*

(Received July 6, 1934)

The fluctuation noises originating in vacuum tubes are treated theoretically under the following headings: (1) *thermal agitation* in the internal plate resistance of the tube, (2) *shot effect and flicker effect* from space current in the presence of space charge, (3) *shot effect* from electrons produced by collision ionization and secondary emission, and (4) *space charge fluctuations* due to positive ions. It is shown that thermal agitation in the plate circuit is the most important factor and should fix the noise level in low noise vacuum tubes; shot noise and flicker noise are very small in tubes where complete temperature saturation is approached; shot noise from secondary electrons is negligible under ordinary conditions; and noise from space charge fluctuation due to positive ions is usually responsible for the difference between thermal noise in the plate circuit and total tube noise. A method is deduced for the *accurate rating of the noise level of tubes* in terms of the input resistance which produces the equivalent thermal noise. *Quantitative noise measurements* by this method are reported on four different types of vacuum tubes which are suitable for use in the initial stage of high gain amplifiers. Under proper operating conditions the noise of these tubes approaches that of thermal agitation in their plate circuits at the higher frequencies and is 0.54 to 2.18×10^{-16} mean square volts per cycle band width in the frequency range from 200 to 15,000 cycles per second. Below 200 cycles per second the noise is somewhat larger. *The minimum noise in different types of vacuum tube circuits* is discussed. These include input circuits for high gain amplifiers, ionization chamber and linear amplifier for detecting corpuscular or electromagnetic radiation, and photoelectric cell and linear amplifier for measuring light signals. With the aid of these results it is possible to design circuits having the maximum signal-to-noise ratio obtainable with the best vacuum tubes now available.

INTRODUCTION

IT is well known that the noise inherent in the first stage of a high gain amplifier is a barrier to the amplification of indefinitely small signals. Even when fluctuations in battery voltages, induction, microphonic effects, poor insulation, and other obvious causes are entirely eliminated, there are two sources of noise which remain, namely, thermal agitation of electricity in the circuits and voltage fluctuations arising from conditions within the vacuum tubes of the amplifier. The effect of thermal agitation in circuits outside the vacuum tube is well understood, but in the case of tube noise there is considerable confusion. In order to clarify the whole subject, the present paper analyzes the various sources of noise in vacuum tubes and their attached circuits, points out a new method for the measurement of tube noise, reports the results of such measurements on four different types of vacuum tubes, and discusses the minimum noise in different types of vacuum tube circuits.

Often, in the use of high gain amplifiers, the impedance of the input circuit is naturally high or may effectively be made high by the use of a transformer. In this case the contribution of noise from the vacuum tube is small compared with the noise arising from thermal agitation in the input circuit. This is a desirable condition since it furnishes the largest ratio of signal to noise for a given input power. Sometimes, however, the input impedance is perforce so small that the tube noise may be comparable with or greater than the thermal agitation noise. Such conditions may arise, for example, in amplifiers where the frequency dealt with is high or the frequency range is wide. It is, therefore, desirable to know the noise level to be expected from different types of tubes that may be used in the first stage of high gain amplifiers as well as to be able to calculate the thermal noise level of the input circuit.

The noise of thermal agitation[1] arises from the fact that the electric charge in a metallic conductor shares the thermal agitation of the molecules of the substance so that minute variations of potential difference are produced between the terminals of the conductor. The mean square potential fluctuation is proportional to the absolute temperature and to the resistive component of

[1] J. B. Johnson, Phys. Rev. **32**, 97 (1928).

the impedance of the conductor, but is independent of the material. The thermal noise power is distributed equally over all frequencies although the apparent magnitude depends on the electrical characteristics of the measuring system as well as on those of the conductor itself. From purely theoretical considerations the following equation has been derived[2] to give the thermal noise voltage at the output of an amplifier due to the thermal agitation of electric charge in an impedance at the input:

$$\overline{E_T^2}=4kT\int_F R(f)\,|G_1(f)|^2 df. \qquad (1)$$

$\overline{E_T^2}$ is here the mean square thermal noise voltage across the measuring device, k is Boltzmann's constant (1.37×10^{-23} watt second per degree), T the temperature of the impedance expressed in degrees Kelvin, $R(f)$ the resistive component of the impedance at the frequency f, $G_1(f)$ the voltage amplification between the input impedance and the measuring device at the frequency f, and F the frequency band within which the amplification is appreciable.

While the thermal noise in the circuit is accurately predictable, the noise originating within the vacuum tube is not completely understood and cannot be calculated accurately. It is known, however, that tube noise arises from a number of different causes, chief among which are: (1) thermal agitation in the internal plate resistance of the tube, (2) shot effect and flicker effect from space current in the presence of space charge, (3) shot effect from electrons produced by collision ionization and secondary emission, and (4) space charge fluctuations due to positive ions. Each of these sources of noise will be discussed in the following section.

Origin of Noise in Thermionic Amplifier Tubes

Thermal agitation in the internal plate resistance of the tube[3]

Just as voltage fluctuations are produced by thermal agitation in resistances comprising the input circuit, so the resistance component of the impedance between plate and cathode is a source of thermal noise. This impedance consists of the internal plate impedance of the vacuum tube in parallel with the external load impedance. Llewellyn[4] has shown that the resistive component of the internal plate impedance produces thermal noise as if it were at the temperature of the cathode. The following formula has been developed by him to cover the case where the tube impedance and load impedance are pure resistances:

$$\overline{E_T^2}=4k[(r_p r_0)/(r_p+r_0)^2](T_0 r_p+T_f r_0)\int_F |G_2(f)|^2 df. \qquad (2)$$

Here r_p is the internal plate resistance of the tube, r_0 the external load resistance in the plate circuit, $G_2(f)$ the voltage amplification between the load resistance r_0 and the measuring device, and T_0 and T_f, respectively, the temperatures of the external load resistance and cathode expressed in degrees Kelvin. The relationship between $G_1(f)$ and $G_2(f)$ is given by

$$G_1(f)=G_2(f)(\mu r_0)/(r_0+r_p), \qquad (3)$$

where μ is the voltage amplification factor of the tube. By assuming $G_2(f)$, in Eq. (2), to be constant over the frequency range F and substituting for it the value given by Eq. (3), it is found on integrating that the thermal noise in the plate circuit of the tube produces the same effect in the measuring device as a signal applied to the input circuit whose magnitude at the grid expressed in mean square volts is given by

$$\overline{V^2}=4kT_0(r_p/\mu)^2[T_f/(T_0 r_p)+1/r_0]F. \qquad (4)$$

Since the noise of thermal agitation is always present, this equation gives the absolute minimum to which fluctuation noise in an amplifying tube can be reduced after all other causes have been eliminated. It shows that for the ideal low

[2] H. Nyquist, Phys. Rev. **32**, 110 (1928).

[3] During the preparation of this paper a paper by E. B. Moullin and H. D. Ellis entitled *Spontaneous Background Noise in Amplifiers Due to Thermal Agitation and Shot Effects* appeared in I.E.E. J. **74**, 323 (1934). The authors there contend that no thermal noise is produced in the plate impedance of a thermionic vacuum tube and that shot noise is not altered by the presence of space charge. With these contentions I cannot agree and I hope to state my definite reasons therefore at a later date.

[4] F. B. Llewellyn, Proc. I.R.E. **18**, 243 (1930).

noise tube in which thermal noise in the plate circuit is the limiting factor, the noise level may be reduced by a decrease in the cathode temperature, a decrease in the effective frequency band, or by an independent decrease in the plate resistance or increase in the amplification factor. In order to operate at a minimum noise level the tube should work into a load resistance which is large in comparison with r_pT_0/T_f. Under this circuit condition the noise level is inversely proportional to μ^2/r_p, a quantity often defined as the "figure of merit" of an amplifying tube.

Shot effect and flicker effect in the presence of space charge

The theory of the shot effect in the absence of space charge has been studied quite completely both theoretically and experimentally by many investigators.[5] The results, however, are not applicable to the study of noise in thermionic vacuum tubes used in high gain amplifiers, since a high degree of space charge is required in tubes used for this purpose. Llewellyn has extended the theory of the shot effect to cases where partial temperature saturation exists, and obtained a general equation to cover all conditions.[4] This equation reduces to the following form when the load impedance is a pure resistance:

$$\overline{E_S^2}=2ej(\partial i/\partial j)^2[r_pr_0/(r_p+r_0)]^2\int_F|G_2(f)|^2df. \quad (5)$$

$\overline{E_S^2}$ is here the mean square shot voltage across the measuring device, i the total space current, j the total current emitted by the cathode, and e the electronic charge (1.59×10^{-19} coulomb).

A precise experimental verification of this equation is very difficult because of the difficulty in determining $\partial i/\partial j$ accurately. Thatcher,[6] however, has made shot measurements in the presence of space charge ($1\geqq\partial i/\partial j\geqq0.66$) which verify the theory within the experimental error of the determination of $\partial i/\partial j$.

Eq. (5) shows that as long as space charge is too small to affect the flow of current, that is when i is equal to j, the mean square shot voltage is directly proportional to the space current. As emission is increased, however, space charge begins to control and finally limits the space current so that the value of $\partial i/\partial j$ approaches zero. Thus the shot voltage increases less rapidly as space charge becomes effective and then finally decreases rapidly toward zero as complete space charge control is reached.

Experimental curves showing the effect of space charge on tube noise are shown in Fig. 1 where abscissae represent space current in milliamperes, and ordinates represent mean square noise voltage across the output measuring device expressed in arbitrary units. The change in space current was obtained by varying the filament heating current while the plate voltage remained constant. Tubes having thoriated tungsten, tungsten, and barium oxide cathodes were used.

At low space currents where no space charge is present the thoriated tungsten and tungsten filaments each give a pure shot effect, the mean square voltage increasing linearly with the space current. As the space current is increased further and space charge sets in, the shot voltage in each

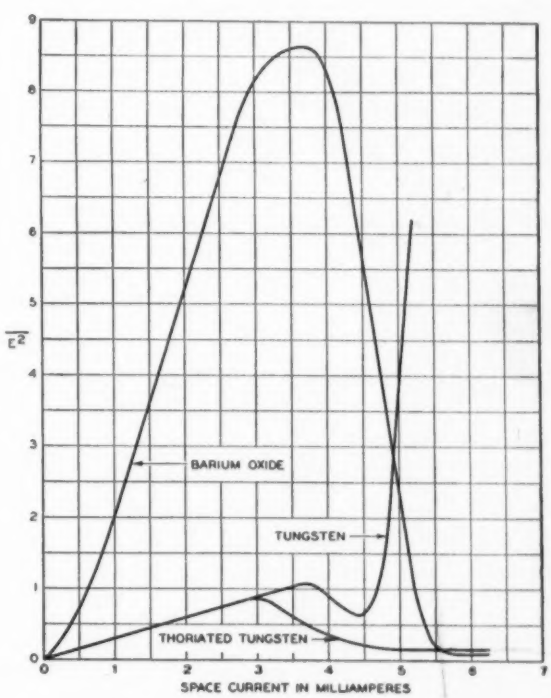

FIG. 1. The effect of space charge on fluctuation noise. Three tubes having filaments composed of tungsten, thoriated tungsten, and barium oxide. $\overline{E^2}$ is the mean-square-noise voltage across the output measuring device expressed in arbitrary units. The variation in space current was obtained by changing the cathode temperature, the plate voltage remaining constant.

[5] W. Schottky, Ann. d. Physik **57**, 541 (1918); T. C. Fry, J. Frank. Inst. **198**, 203 (1925); A. W. Hull and N. H. Williams, Phys. Rev. **25**, 147 (1925).

[6] Everett W. Thatcher, Phys. Rev. **40**, 114 (1932).

tube goes through a maximum and decreases with oncoming temperature saturation as suggested by Eq. (5). With the approach of complete temperature saturation the noise, however, does not decrease to zero in accordance with this equation. If it were possible to reach complete temperature saturation the residual noise would not be due to the shot effect, but rather to thermal noise in the plate circuit of the tube, positive ions and secondary emission within the tube, and other contributing causes. Usually this condition is approached in the better commercial tubes so that the contribution of true shot noise is a small part of the total noise.

If the methods used in obtaining Eq. (4) are applied to Eq. (5), it is found that the shot noise in the plate circuit of the tube produces the same effect in the output measuring device as a signal applied to the input circuit whose magnitude at the grid expressed in mean square volts is

$$\overline{V^2}=2ej(\partial i/\partial j)^2(r_p/\mu)^2F. \qquad (6)$$

This equation shows that the level of shot noise at the input is lowered by an increase in the cathode temperature, which increases the degree of temperature saturation, and by an increase in the ratio μ/r_p, which by definition is the transconductance of the tube, but is independent of the external load resistance. It should be remembered, however, that shot noise in the plate circuit should not fix the noise level in low noise vacuum tubes and that never, as is sometimes done in the literature, can the noise of an amplifier be calculated as pure shot noise in the plate circuit, for in the absence of space charge the tube would not be an amplifier.

Although space charge can counteract the effect of random electron emission from the cathode so that shot noise is reduced, other factors can alter the flow of current in such a way that the noise is increased. This is the case when changes in emission occur over small areas of the cathode, giving rise to an additional fluctuation which has been termed flicker effect.[7] This type of noise is particularly noticeable with oxide-coated cathodes. Since the flicker effect is due to localized variations in the emission of the cathode, one would expect it to disappear in the presence of a complete space charge condition.

The experimental curve for the barium oxide coated filament, Fig. 1, shows a flicker effect many times larger than the shot effect on which it is superimposed. At low space currents the mean square flicker effect voltage increases faster than the pure shot noise, a square law rather than a linear relationship being followed. As space charge sets in, the flicker effect voltage goes through a maximum and then decreases with increased space current in the same manner as does the shot effect voltage. In spite of the large flicker effect, as complete temperature saturation is approached the total noise is even less than that found with the thoriated tungsten filament which has no flicker effect. This illustrates clearly the effectiveness of space charge in smoothing the space current.

When the control grid of a vacuum tube is floating at its equilibrium potential, the noise level is much higher than when the grid is connected through an input circuit to the cathode. This increase in noise is primarily due to thermal noise in the extremely high input resistance of the tube and to shot noise arising from small grid currents.[8] The magnitude of the thermal noise may be calculated knowing that the input impedance of the tube consists of its input resistance, r_g, in parallel with its dynamic grid-to-ground capacity. In such a combination the real resistance component, $R(f)$, is related to the pure resistance, r_g, and the dynamic capacity, c, according to the equation

$$R(f)=r_g/(1+4\pi^2c^2r_g^2f^2). \qquad (7)$$

According to Eq. (1) the mean square thermal noise input voltage is then

$$\overline{V_T^2}=4kTr_g\int_F df/(1+4\pi^2c^2r_g^2f^2). \qquad (8)$$

With the grid floating at its equilibrium position (usually slightly negative with respect to the cathode) the grid current is composed of two components equal in magnitude but opposite in sign. The one component consists of electrons reaching the grid, while the other consists of positive ions reaching and electrons leaving the grid. The electrons are liberated from the grid by

[7] J. B. Johnson, Phys. Rev. **26**, 71 (1925); W. Schottky, Phys. Rev. **28**, 74 (1926).

[8] L. R. Hafstad, Phys. Rev. **44**, 201 (1933).

secondary emission, the photoelectric effect, thermionic emission, and soft x-rays. It should be pointed out that space charge does not reduce the noise produced by the shot effect in any of these currents. The general shot effect equations[9] show that the magnitude of shot noise from these grid currents is

$$\overline{V_s^2} = 2ei_g r_g^2 \int_F df/(1+4\pi^2 c^2 r_g^2 f^2)^2, \qquad (9)$$

where i_g is the sum of the grid currents regardless of sign.

Noise produced by secondary effects

In this classification are grouped several sources of disturbance whose individual effects are very difficult to calculate and measure under the operating conditions of the vacuum tube. For this reason the following discussion will include only a general consideration of the more obvious contributing causes.

Although the cathode is the principal source of electrons which reach the plate, in actual practice electrons are produced by ionization of the gas molecules within the tube or by secondary emission resulting from bombardment of the tube elements. Electrons produced in this manner are drawn to the plate and generate noise which is not much affected by the space charge. Assuming a reasonable magnitude for the current produced in this manner it can be shown by the shot equations that noise from this source is usually negligible. In cases where the gas pressure within a tube is above normal, or in screen-grid and multi-grid tubes having high plate resistances and considerable secondary emission, the shot noise from secondary and ionization electrons may be of the same order of magnitude as thermal noise in the plate circuit.

Positive ions formed from ionized gas molecules or emitted from the tube elements are much more effective in producing noise since, instead of being drawn off to the plate, they are attracted into the space charge region where small disturbances in equilibrium produce large momentary fluctuations in space current. Due to their large mass the motions of the ions are relatively slow, so that they are very effective in this respect. This type of noise is quite disturbing in amplifying tubes for it tends to become a maximum at complete temperature saturation. This is illustrated very clearly in the noise measurements on the tungsten filament shown in Fig. 1. Here positive ions from the filament begin to show their effect as space charge sets in, the number of ions and the amount of noise increasing as temperature saturation is approached. As heard in the loudspeaker, this noise consists of sharp crackling sounds which can easily be distinguished from the steady rustling noise of the shot and thermal effects.

Ballantine[10] has recently made calculations and measurements on the noise due to positive ions from collision ionization in which he has shown that the mean square noise voltage is roughly proportional to the gas pressure within the tube and to the 3/2 power of the plate current. Comparing his results with Eq. (2), it appears that under ordinary working conditions the noise due to collision ionization in a vacuum tube may be of the same order of magnitude as noise from thermal agitation in its plate circuit. The noise level of tubes having a poor vacuum, however, may be much higher.

Measurement of Tube Noise

The performance, as regards freedom from noise, of a vacuum tube used in an amplifier may be indicated by a comparison between the noise and a signal applied to the grid. Usually we say that the noise is equivalent to a signal which gives the same power dissipation in the output measuring instrument as the noise, the frequency of the signal being suitably chosen with respect to the frequency characteristics of the amplifier. Since tube noise is distributed over all frequencies and the noise power increases with the effective band width, it will be advantageous to express this input signal in equivalent mean square volts per unit frequency band width, effective over a given frequency range.

From these considerations it can be seen that the most convenient standard signal for measuring the equivalent input noise over any given frequency range is one in which the mean square signal voltage is distributed equally over all frequencies. With such a signal the equivalent in-

[9] E.g., reference (4) or (5).

[10] Stuart Ballantine, Physics 4, 294 (1933).

put noise over any frequency range can be measured directly, while if an oscillator is used a number of measurements are required and the result must be computed by graphical integration. A signal which meets these frequency requirements perfectly is the noise of thermal agitation. Accordingly, in the measurements to be described here the standard input signal will be the thermal agitation voltage of a resistance R connected between the control grid and cathode of the tube under test.[1]

The thermal noise voltage of the grid circuit, referred to the output measuring device, is given by Eq. (1), where $R(f)$ is the real resistance component of an input impedance consisting of the pure resistance R in parallel with its shunt capacity and that of its leads and of the vacuum tube. In such a combination $R(f)$ is related to the pure resistance R and the total capacity c according to Eq. (7). In all the measurements described here the factor $4\pi^2c^2R^2f^2$ is so small in comparison with unity that it may be neglected without appreciable error. Under these conditions Eq. (1) reduces to

$$\overline{E_T^2}=4kTR\int_F |G_1(f)|^2df, \qquad (10)$$

where R is the direct-current value of the resistance between control grid and cathode of the tube under test.

The voltage fluctuations arising from conditions within the tube produce a mean square voltage output $\overline{E_N^2}$ according to the equation

$$\overline{E_N^2}=\int_F |V(f)|^2|G_1(f)|^2df, \qquad (11)$$

where $|V(f)|^2$ is the tube noise at the frequency f for unit frequency band width, expressed in volts squared and referred to the input circuit. Letting V_F^2 be the effective value of $|V(f)|^2$ over the band width of the amplifier we obtain

$$\overline{E_N^2}=V_F^2\int_F |G_1(f)|^2df. \qquad (12)$$

Since the integrals in Eqs. (10) and (12) are identical it is found on dividing one equation by the other and solving for V_F^2 that:

$$V_F^2=4kTR(\overline{E_N^2}/\overline{E_T^2}). \qquad (13)$$

Eq. (13) enables one to calculate the magnitude of tube noise in the frequency range F, per unit cycle band width, in terms of the thermal noise generated in a resistance R placed in the input circuit.[11] Since this equation contains no integral the measurements are simplified in that neither standard signal generator nor calibrated amplifier are required.

Apparatus

The experimental arrangement used in the measurements to be reported here is given in schematic form in Fig. 2. The system includes the tube under test, a high gain amplifier, appropriate filters, an attenuator, and an output measuring device.

The input circuit consists of the tube under test together with the variable grid resistor, external load resistor, and batteries for furnishing the required filament, grid and plate voltages. Because of the high value of amplification required and the wide frequency range covered by the amplifier, this circuit required shielding from external disturbances arising from electrical, mechanical and acoustical shock. Accordingly the tube under test was suspended by means of rubber bands, the whole circuit with the exception of batteries placed inside a tightly sealed lead lined box, and this box in turn suspended by means of a system of damped springs. The box with its cover removed and the tube in place is shown in Fig. 3. This shielding was sufficient to reduce the noise from outside disturbances to such a low level that no correction had to be made for it at any time.

The high gain amplifier[12] consists of two separate resistance coupled units each containing

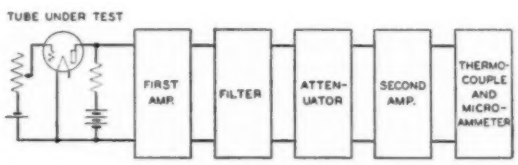

Fig. 2. Schematic amplifier circuit for measuring fluctuation noise in vacuum tubes.

[11] It is assumed that tube noise does not vary with frequency, or that the band width of the amplifier is so narrow that no appreciable error is introduced in applying the result.

[12] The essential parts of this amplifier were designed by Mr. E. T. Burton.

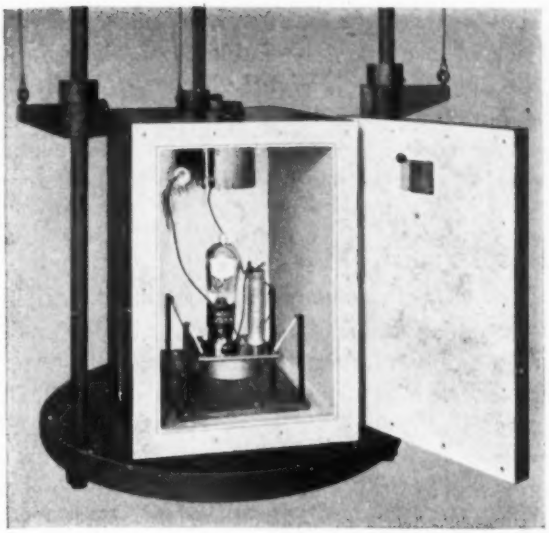

FIG. 3. Tube under test mounted in the shielding box.

three stages. Each unit is so designed and shielded that the effect of external disturbances is eliminated. The total gain obtainable is about 164 db (constant to within 2 db from 10 cycles to 15,000 cycles). Since this gain is in excess of that required for the study of thermal and tube noises, an attenuator having a range of 63 db was inserted between the two units. In order to limit amplification to certain desired frequency bands, specially designed electric filters were inserted between the first amplifier unit and the attenuator. Three such filters were used of which one is a low pass filter with cut-off around 205 cycles, and the other two are band pass filters with mid-frequencies at 1750 and 11,000 cycles, respectively. The frequency characteristic of the amplifier with no filter and with each filter inserted is shown in Fig. 4.

The recording instrument is a 600-ohm vacuum thermocouple and microammeter. Conveniently, the deflection of the microammeter is closely proportional to the mean square voltage applied to the couple. The procedure in making a measurement of tube noise is as follows: With the tube under test operating at zero grid resistance, the attenuator is adjusted to give a convenient deflection of the microammeter (due to noise in the tube under test). Grid resistance is now added until this deflection is exactly doubled thus making $\overline{E_N^2}$ equal to $\overline{E_T^2}$. This value of input

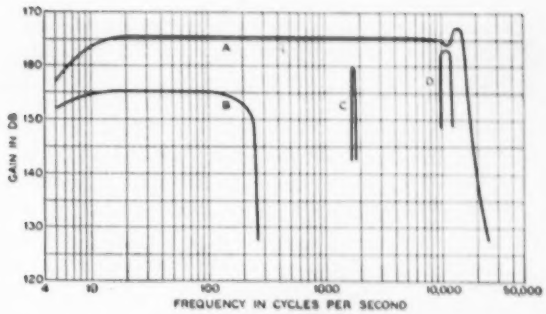

FIG. 4. Frequency characteristic of amplifier circuit. Curve A with no filter, curve B with low pass filter, and curves C and D with band pass filters.

resistance, designated by R_G, is a measure of the inherent noise of the tube. Substituting R_G in Eq. (13) the tube noise is calculated from the relation

$$V_F^2 = 4kTR_G = 1.64 \times 10^{-20} R_G \text{ (volt)}^2, \quad (14)$$

where R_G is expressed in ohms and T is 300 degrees Kelvin (approximate room temperature).

NOISE IN CERTAIN VACUUM TUBES

Quantitative measurements of tube noise were made on four different types of standard Western Electric vacuum tubes, namely: Nos. 102G, 264B, 262A and 259B. These tubes have as low a noise as any tube obtainable at the present time.[13]

In order to obtain the best signal to noise ratios it was found that operating conditions different from those normally recommended must be used. In general, the cathode must be operated at as high a temperature as possible without impairing the life of the tube, the negative bias of the control grid must be reduced to as near zero as possible without causing excessive grid current, and the plate voltage must be reduced below the value normally recommended. In all the measurements described here the tube under test was coupled to the first amplifier unit through a 50,000-ohm load resistance. It was found that the signal-to-noise ratio could be improved a fraction of a db by increasing the load resistance (in accordance with Eq. (4)); this, however, necessitated a large plate voltage which was in-

[13] Noise measurements on R.C.A. 38 and General Electric P.J. 11 tubes have recently been reported by E. A. Johnson and C. Neitzert, R.S.I. 5, 196 (1934).

convenient. Six tubes of each type were tested and the noise data given below were obtained by averaging the six measurements for each type. Individual tubes may differ from these average values by as much as ±1 db.

No. 102G tube

This is a three-element, filament-type tube. Its long life, exceptionally high stability of operation, and good temperature saturation make it a desirable tube to use in the input stage of certain high gain amplifiers. This tube also has a comparatively small microphonic response to mechanical and acoustical shock although it is not as good as the No. 262A and the No. 264B tubes in this respect.

The conditions found most suitable for quiet operation of the No. 102G tube and the corresponding average tube characteristics are given in the first two columns of Table I. Under these conditions the average equivalent tube noise voltage, referred to the grid circuit, is given in the last column of the same table. These noise data are given in terms of R_G, the experimentally determined equivalent noise resistance of the tube, and in terms of V_F^2, calculated by means of Eq. (14), for each of the four frequency ranges shown in Fig. 4.

The No. 102G has the lowest noise of all the tubes tested and was found suitable for use in the first stage of high gain amplifiers where tube noise is the limiting factor, providing it is not required that the input capacity and microphonic response to mechanical and acoustical shock be extremely low.

No. 264B tube

This is a three-element filament-type tube. Due to the rigid construction and the short filament which is designed to reduce vibration to a minimum, the microphonic response of the tube to mechanical and acoustical shock is exceptionally low.[14] The extensive system of spring suspensions and the heavy sound-proof chamber usually required for shielding low noise tubes may be simplified when using the No. 264B. In addition, this tube has good temperature saturation, low power consumption, and high stability of operation.

The operating conditions and noise data for this tube are given in Table II. Although the noise of this tube is slightly higher than that of the No. 102G, the lower microphonic response and the lower power consumption make it a more desirable tube to use in input stages of certain high gain amplifiers.

No. 262A tube

This is a three-element tube having an indirectly heated cathode. It is designed to give a microphonic response to mechanical and acoustical shock[14] still lower than that of the 264B. Except for frequencies below 200 cycles per second it was found that no acoustic shield was necessary for this tube even when working at extremely low levels. Although this tube is designed to have a low hum disturbance resulting from alternating current for heating the cathode, (the interference from this effect can be held to less than 7×10^{-6} equivalent input volt) direct-current power was used in the measurements here described.

The operating conditions and noise data for the No. 262A tube are given in Table III.

No. 259B tube

This is a four-element, screen-grid tube having an indirectly heated cathode. Its comparatively high amplification factor makes possible a relatively large gain per stage so that when it is used in the first stage of a high gain amplifier succeeding stages contribute nothing to the total noise.

Noise measurements on the No. 259B tube show that the signal-to-noise ratio is approximately independent of the plate voltage over a wide operating range, but is closely dependent on the plate current as affected by the control and screen grid voltages. Table IV contains the operating conditions and noise data for this tube.

Noise measurements were also made on the No. 259B tube with its control grid floating at equilibrium potential. Using the operating voltages specified above, the noise level was about 20 db higher than those given in Table IV. The level can be greatly reduced by operating the tube at a lower cathode temperature and with lower screen and plate voltages.[15] This reduction in noise is due to a decrease in current to the floating grid.

[14] M. J. Kelly, Soc. Motion Picture Eng. J. **18**, 761 (1932).

[15] I am indebted to Dr. J. R. Dunning of Columbia University for pointing out this fact.

TABLE I. *Western Electric No. 102G tube.*

Operating Conditions		Tube Characteristics		Noise Data: Frequency range (cycles per sec.)	Noise Data: V^2_F (volt)^2	Noise Data: R_G (ohms)
Filament						
Voltage	2.0 v	Type of tube	3 element			
Current	1.0 amp.	Type of cathode	Oxide-coated filament	10–15,000	0.64×10^{-16}	3,900
Grid						
Voltage	−0.5 v	Amplification factor	30	5–205	2.2	13,600
Plate						
Voltage	130 v	Plate resistance	45,000 ohms	1,750–1,850	0.58	3,550
Current	1.2 m.a.					
Load						
Resistance	50,000 ohms	Approx. dynamic input capacity	80 $\mu\mu$f	10,000–12,000	0.54	3,300

TABLE II. *Western Electric No. 264B tube.*

Operating Conditions		Tube Characteristics		Noise Data: Frequency range (cycles per sec.)	Noise Data: V^2_F (volt)^2	Noise Data: R_G (ohms)
Filament						
Voltage	1.5 v	Type of tube	3 element			
Current	.30 amp.	Type of cathode	Oxide-coated filament	10–15,000	1.3×10^{-16}	7,650
Grid						
Voltage	−0.5 v	Amplification factor	7	5–205	6.6	40,000
Plate						
Voltage	26 v	Plate resistance	18,500 ohms	1,750–1,850	1.1	6,800
Current	0.6 m.a.					
Load						
Resistance	50,000 ohms	Approx. dynamic input capacity	30$\mu\mu$f	10,000–12,000	1.0	6,200

TABLE III. *Western Electric No. 262A tube.*

Operating Conditions		Tube Characteristics		Noise Data: Frequency range (cycles per sec.)	Noise Data: V^2_F (volt)^2	Noise Data: R_G (ohms)
Heater						
Voltage	10 v	Type of tube	3 element			
Current	0.32 amp.	Type of cathode	Oxide-coated indirectly heated	10–15,000	1.3×10^{-16}	7,700
Grid						
Voltage	−1.0 v	Amplification factor	15.7	5–205	17.	100,000
Plate						
Voltage	44 v					
Current	1.0 m.a.	Plate resistance	22,000 ohms	1,750–1,850	1.0	6,400
Load						
Resistance	50,000 ohms	Approx. dynamic input capacity	23 $\mu\mu$f	10,000–12,000	0.84	5,100

TABLE IV. *Western Electric No. 259B tube.*

Operating Conditions		Tube Characteristics		Noise Data: Frequency range (cycles per sec.)	Noise Data: V^2_F (volt)^2	Noise Data: R_G (ohms)
Heater						
Voltage	2.0 v	Type of tube	4 element screen grid			
Current	1.7 amp.	Type of cathode	Oxide-coated indirectly heated	10–15,000	3.2×10^{-16}	19,800
Grid						
Control voltage	−1.5 v	Amplification factor	1,500	5–205	7.7	47,000
Screen voltage	22.5 v					
Plate						
Voltage	100 v	Plate resistance	2.75 megohms	1,750–1,850	2.8	17,100
Current	0.6 m.a.					
Load						
Resistance	50,000 ohms	Approx. dynamic input capacity	6.0 $\mu\mu$f	10,000–12,000	2.8	17,000

By using a heater current of 1.3 amperes, a plate current of 0.1 milliampere, a screen potential of 16.5 volts and a plate potential of 30 volts the equivalent input noise was 1.4×10^{-5} volt for the entire frequency range from 10 cycles to 15,000 cycles. Under these operating conditions the floating grid potential was 1.0 volt negative with respect to the cathode, the input resistance 1.4×10^{10} ohms, the dynamic grid-to-cathode capacity 6×10^{-12} farad, and each component of grid current about 4.5×10^{-12} ampere.

Discussion of Results

From the noise data in Tables I–IV one can estimate quite accurately the equivalent input noise voltage of each of the four types of tubes at any frequency between 5 and 15,000 cycles, and for any band width within these limits. For example, using the noise data given in Table I the equivalent input noise voltage of the No. 102G tube working over a band having sharp cut-offs at 5 cycles and 205 cycles is computed to be

$$(\overline{V^2})^{\frac{1}{2}}=(V_F^2F)^{\frac{1}{2}}=2.1\times10^{-7}\text{ volt.} \qquad (15)$$

For a band width of 200 cycles with mid-frequency at 10,000 cycles this noise is reduced to 1.0×10^{-7} volt. It can be seen that for each type of tube the noise voltage over equal band widths is between 1.5 and 4.5 times greater at frequencies below 200 cycles than at the higher frequencies.[16]

Even at high frequencies the noise voltage is above that expected from thermal noise in the plate circuit which, as stated above, is the absolute minimum to which fluctuation noise in a thermionic vacuum tube may be reduced after all other causes are eliminated. In the case of the No. 102G tubes for instance, by use of the operating conditions of Table I, and assuming 1100 degrees Kelvin as the temperature of the barium oxide filament, it is found by means of Eq. (4) that the equivalent input noise voltage produced by thermal agitation in the plate circuit is 2.7×10^{-8} volt for a band width of 200 cycles. The total input noise voltage obtained experimentally at the higher frequencies is greater than this by a factor of 3.8. In like manner it is found that the total input noise voltages found experimentally for the Nos. 264B, 262A and 259B tubes are greater than the equivalent input thermal noise voltages produced in the plate circuit by factors 2.1, 3.7 and 16, respectively. These calculations show that each of these four types of tubes approaches the requirements of an ideal low noise amplifying tube although none of them is perfect in this respect.

As stated above, the best signal-to-noise ratio in a high gain amplifier is obtained when thermal agitation in the input resistance is responsible for most of the noise in the amplifier. This condition is met when the resistance of the input circuit is higher than the value of R_G for the input tube. In case the resistance in the input circuit is less than R_G the input signal and the thermal noise from the input circuit can be raised above the noise of the tube by using an input transformer having a sufficiently high voltage step-up. The voltage ratio of the transformer, and in turn the possible ratio of input circuit thermal noise to tube noise, is limited, especially at the higher frequencies, by the dynamic grid-to-ground capacity of the input tube and its leads. In such a circuit the No. 259B tube with its lower interelectrode capacities and higher tube noise is often more desirable than even the quietest three-element tubes.

In those high gain amplifiers in which unavoidably the resistance of the input circuit is low, the tube rather than thermal agitation in the grid circuit is responsible for most of the noise. Here the best signal-to-noise ratio can be obtained by choosing a tube for the initial stage having the lowest possible noise level. The above measurements show that one of the three element tubes, particularly the No. 102G tube if sufficient shielding is used, is best suited for this purpose.

The lower limits of noise obtainable with high gain amplifiers may be estimated by means of Fig. 5, which shows the noise as a function of input resistance and frequency band width when thermal agitation in the input circuit is responsible for all the noise. The data for this figure are obtained from the thermal noise relationship

$$\overline{V_T^2}=1.64\times10^{-20}RF\text{ (volt)}^2. \qquad (16)$$

R is expressed in ohms and the temperature has been taken at 300 degrees Kelvin which is ap-

[16] Other investigators have also found an increase in tube noise energy at the lower frequencies. G. F. Metcalf and T. M. Dickinson, Physics **3**, 11 (1932).

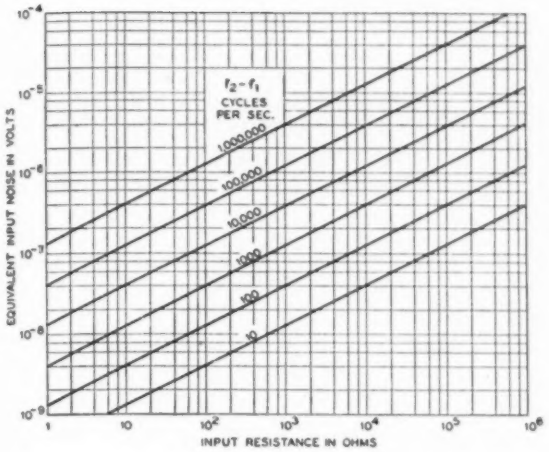

FIG. 5. Thermal noise level as a function of input resistance and frequency range.

proximately room temperature. It must be remembered that the attainment of these noise levels at low input resistances is limited by the input transformer.

The results of the noise measurements on the No. 259B tube with floating grid may be compared with the value predicted by Eqs. (8) and (9). Inserting the tube characteristics obtained by experiment ($r_g = 1.4\times10^{10}$ ohms, $i_g = 9\times10^{-12}$ ampere, and $c = 6\times10^{-12}$ farad), and integrating between the frequency limits 10 cycles and 15,000 cycles, it is found that the equivalent thermal noise input is 0.9×10^{-5} volt, while the shot noise input is 1.4×10^{-5} volt. The total noise is the square root of the sum of the squares of these values or 1.7×10^{-5} volt. This agrees with the measured value of 1.4×10^{-5} volt within an error of 20 percent, which is as accurate as the determination of the grid currents. These equations may also be used to calculate the noise originating in the grid circuit when external resistance or capacity is connected between grid and cathode, providing r_g and c are now calculated from the internal and external impedances in parallel.

A common method of detecting corpuscular or electromagnetic radiation makes use of an ionization chamber and linear amplifier. In this circuit the control grid in the first tube of the amplifier is connected to the collecting electrode of the ionization chamber and both allowed to float at equilibrium potential.[17] The shot and thermal noise in this grid circuit sets a limit to the measurement of extremely weak radiation. Knowing the value of input capacity, input resistance, floating grid current, and the frequency limits of the amplifier, Eqs. (8) and (9) may be used to calculate this limiting noise level. For example, if one uses a No. 259B tube with the operating voltages specified for floating grid, an ionization chamber having a capacity of 15×10^{-12} farad, and an amplifier having a frequency range from 200 to 5000 cycles per second the limiting noise level is 1×10^{-6} root-mean-square volt.

The limiting noise level in a sytem consisting of a photoelectric cell and thermionic amplifier is determined by thermal agitation in the coupling circuit between the photoelectric cell and amplifier, and by shot noise in the photoelectric current (in circuits where the photoelectric current is very small and the coupling resistance is very high, shot noise from grid current in the vacuum tube becomes appreciable). The noise of thermal agitation may be calculated by means of Eq. (8) providing r_g is now replaced by R, the coupling resistance. If vacuum cells are used, the photoelectric current produces a pure shot noise which can be calculated by Eq. (9) providing i_g is replaced by I, the photoelectric current. In gas-filled photo-cells where collision ionization occurs, the noise is in excess of the value calculated in this manner.[18] The relative magnitude of shot noise and thermal noise depends on the values of I and R, and by combining Eqs. (8) and (9) it is found that

$$\overline{V_S{}^2}/\overline{V_T{}^2} = eIR/2\,kT = 19.4\,IR, \qquad (17)$$

where I is expressed in amperes, R in ohms, and T is 300 degrees Kelvin. Thus an increase in either I or R will tend to make shot noise exceed thermal noise. This is the desirable condition since it furnishes the largest ratio of signal-to-noise for a given light signal on the photoelectric cell.

In conclusion I wish to acknowledge my indebtedness to Dr. J. B. Johnson for the helpful criticism he has given during the course of this work.

[17] H. Greinacher, Zeits. f. Physik **36**, 364 (1926).
[18] B. A. Kingsbury, Phys. Rev. **38**, 1458 (1931).

The Behavior of a Copper Oxide Photoelectrolytic Cell When Used with Sinusoidally Varying Illumination*

WILBUR E. MESERVE, *Department of Physics, Cornell University*

(Received June 11, 1934)

Each of the copper electrodes of the cell was coated chemically with a layer of cuprous oxide. The electrolyte used was a one percent solution of lead nitrate. With an amplifier and cathode-ray oscilloscope the open circuit e.m.f. of the cell was studied for sinusoidally varying illumination of one electrode. The frequency response, the spectral response, and the phase shift between e.m.f. and illumination were determined. A.c. bridge measurements showed that the capacitance of the cell was large and that the greater portion of it conforms to the relation for electrolytic capacitance, $C = C_0/f^{\frac{1}{2}}$. The capacitance increases with illumination. The results are interpreted in terms of an equivalent electrical network whose elements have values corresponding to those of the cell as measured with the bridge. The computed results for the network are in excellent agreement with the observed results for the cell throughout the whole frequency range (0 to 10,000) when it is assumed, (1) that the primary effect of illumination is to cause electrons to pass through the "barrier layer" from the oxide to the copper, and, (2) that the number of such electrons is linearly proportional to the intensity of the light. The results thus lead to the conclusion that the photoactivity of this cell is to be explained in the same manner as that of the dry cuprous oxide cell, except that the electrolyte takes the place of the front electrode in making contact with the oxide.

ELECTROLYTIC cells in which the electrodes are of copper covered with a layer of cuprous oxide have long been known to be light sensitive.[1] If one electrode of such a cell is exposed to light, while the other is kept in the dark, an e.m.f. is developed which tends to cause current to flow through the liquid from the dark electrode to the electrode that is illuminated. The magnitude of the current has been found to depend upon the nature of the electrolyte,[2] the thickness of the layer of oxide,[2, 3] and the wave-length of the exciting light.[4] But while cells of this type have been extensively studied, no satisfactory explanation of their behavior has yet been developed. Some regard the effect as fundamentally photoelectric, but most writers look upon it as photochemical in character. It has been suggested that the photovoltaic sensitivity is connected in some way with the effect of light on the electrical resistance of cuprous oxide, or with the change reported by Kennard and Dieterich[5] in the contact potential.

A few years ago it was found that a layer of cuprous oxide on copper may be used to make a photoactive cell even when no liquid electrolyte is present.[6] In these dry copper oxide cells the copper plate serves as one terminal, as in the wet cell, while for the second terminal a semi-transparent layer of metal deposited on the Cu_2O is generally employed. Cells of this type have recently been extensively studied,[7, 8] not only from a scientific standpoint, but as practical devices for the measurement of light intensities.

There can be little doubt that the photoactivity of the two types of cell depends, at least in part, upon phenomena that are common to both. In view of the conflicting theories concerning the effect, it has therefore seemed desirable to undertake a study of the behavior of the wet cell by a method which is radically different from those previously used, and which seemed likely to bring out both the similarities and the differences in the behavior of the two types of cell.

* A brief account of the work here described was presented at the New Haven meeting of the American Physical Society, June, 1932. See W. E. Meserve, Phys. Rev. **41**, 396 (1932).

[1] W. Hankel, Wied. Ann. **1**, 402 (1877).
[2] A. D. Garrison, J. Phys. Chem. **27**, 607 (1923).
[3] T. W. Case, Trans. Am. Electrochem. Soc. **31**, 351 (1917).
[4] V. P. Barton, Phys. Rev. **23**, 337 (1924).
[5] E. H. Kennard and E. O. Dieterich, Phys. Rev. **9**, 58 (1917).
[6] L. O. Grondahl, Science **36**, 306 (1926).
[7] W. Schottky, Phys. Zeits. **31**, 913 (1930).
[8] O. V. Auwers and H. Kerschbaum, Ann. d. Physik **7**, 129 (1930).

Method

The cell used consisted of two electrodes of copper, coated with a thin layer of cuprous oxide. The electrolyte was a one percent solution of lead nitrate, this being selected because it was felt that with the use of a salt that is inert with respect to the copper and the oxide many disturbing and masking chemical phenomena would be eliminated. The cuprous oxide surface was prepared chemically in the manner described by Garrison.[2] All parts of the electrodes not exposed to the light were covered first with opaque lacquer and then with paraffin in order to reduce to a minimum the leakage paths from the illuminated to the unilluminated portions of the same electrode.

All measurements were made while the cell was exposed to light whose intensity varied sinusoidally. In Fig. 1, the upper diagram shows

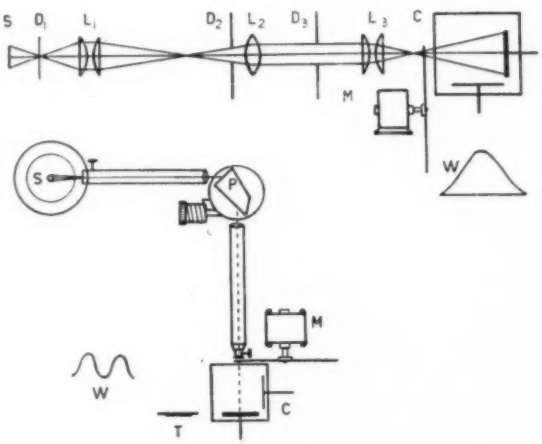

FIG. 1. Upper diagram; arrangement for use with white light. *W*, shape of stationary slot D_3. *M*, variable speed motor. Lower diagram; arrangement for use with monochromatic light. *W*, shape of periphery of aluminum disk attached to motor *M*. *T*, Coblentz thermopile.

the arrangement of apparatus when using white light for determining the frequency response and phase shift of the cell, while the lower diagram is the arrangement for determining its spectral sensitivity. In the upper diagram, *S* is a 500 watt projection lamp operated at constant voltage. D_1 and D_2 are diaphragms having circular openings, while diaphragm D_3 has an opening of sine squared form as shown at *W*. The periphery of the disk attached to the motor *M* was cut in the form of rectangular slots of such dimensions that their width was equal to the width of the image produced by *W*, the disk being in the same plane as the image of *W*. This arrangement gave a sinusoidal variation of the modulated light flux[9] of the type, $\phi = \Phi(1+\sin \omega t)$.

The a.c. component of the potential produced by the varying illumination was amplified by a calibrated vacuum tube amplifier[10] and was then impressed on a Bedell-Reich stabilized cathode-ray oscilloscope.[11] The cell was connected to the grid of the first tube through a 0.5μf condenser with no grid leak resistance (floating grid). Thus only the a.c. component of the photo-potential was amplified and the magnitude of this potential was determined by measuring the amplitude of the deflection of the cathode spot on the fluorescent screen of the cathode-ray tube. The frequency of the a.c. potential produced with the rotating disk method of light modulation was determined by the beat method, using the oscilloscope as an indicator and a General Radio, Type 377-B, sine wave audio oscillator as the frequency standard. This method gave far more accurate values of frequency than could be obtained from measurements of the speed of the motor.

Because of the limitations of the motor speed, the maximum frequency of modulated light flux obtainable was about 6000 cycles per second. In order to extend the response curves to 10,000 cycles, an Aeo glow lamp[12] of the type used in the recording of sound on film was employed. The light output of this lamp, when operated on about 250 volts d.c. (which is well above the ignition potential) is modulated in accordance with any variations in current which may be superimposed upon the applied d.c. voltage. A sinusoidal modulation of light flux was obtained from this lamp by impressing the output of the general radio oscillator on the glow lamp circuit.

[9] For details of this method see, Standards of Inst. Radio Eng., Yearbook, 1931, p. 174.

[10] For a detailed description of the amplifier, its calibration, and its use in connection with photoelectric cells and the oscilloscope, see: *The Measurement and Recording of Instantaneous Light Variations by Means of the Photoelectric Cell and Stabilized Oscilloscope*, W. E. Meserve, Trans. Ill. Eng. Soc. **24**, 671 (1929); also, *A Method for Measuring the Amplitude and Frequency of Vibration of Bodies which may be Heated to Incandescence*, W. E. Meserve, R.S.I. **2**, 47 (1931).

[11] F. Bedell and H. J. Reich, J. A.I.E.E. **46**, 563 (1927).

[12] T. W. Case, U. S. Patents Nos. 1,816,825; 1,822,865; 1,923,051, and others.

The lamp was first calibrated so that fluctuating light of the same peak value could be obtained for all frequencies of impressed a.c. voltage. This was accomplished by using a Burt high vacuum photoelectric cell in place of the cuprous oxide cell and determining the a.c. oscillator input to the glow lamp necessary to give the same deflection of the cathode spot for all frequencies. The fact that the light flux was varying sinusoidally was also determined with the Burt cell by observing the wave form of the Burt cell output simultaneously on the oscilloscope screen with the wave form of the oscillator output voltage, the effect being obtained by means of revolving switch contacts.[11] By means of this visual test no appreciable difference between the shape of the two waves could be detected.

Referring to the lower diagram of Fig. 1, S is the source of light associated with the spectrometer (denoted by P in its prism). The cell C may be replaced by the Coblentz thermopile T. Between the cell and the source is situated a revolving disk attached to the variable speed motor M. The periphery of the aluminum disk is plotted according to the relation $y = A \sin^2 x$, as shown at W. By means of this arrangement the amplitude of the long narrow strip of monochromatic light from the spectroscope was varied sinusoidally, which resulted in a sinusoidal variation of the total energy of the incident light and a consequent sinusoidal pulsating photo-potential.

Results

Fig. 2 shows the frequency response curves of the cell for the four relative intensities indicated. For frequencies not exceeding 500 cycles per sec.

TABLE I. *The photo-e.m.f. multiplied by relative intensity is a constant for each frequency.*

Freq. (cycles/ sec.)	Curve (1) Photo-e.m.f. ×3.0	Curve (2) Photo-e.m.f. ×2.0	Curve (3) Photo-e.m.f. ×1.5	Curve (4) Photo-e.m.f. ×1.0
	($\times 10^{-2}$)	($\times 10^{-2}$)	($\times 10^{-2}$)	($\times 10^{-2}$)
20	6.3	6.28	6.3	6.2
50	6.6	6.6	6.7	6.5
100	6.45	6.5	6.6	6.5
200	5.55	5.4	5.45	5.5
500	3.8	3.9	3.9	4.1
1000	2.4	2.0	1.8	1.9

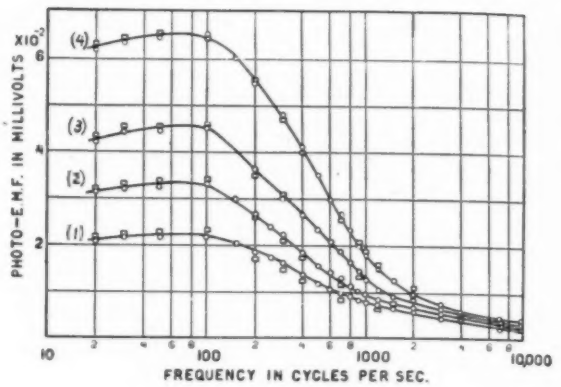

FIG. 2. A.c. photo-potential-light modulation frequency variation for the four following relative illumination intensities. Curve (1), relative intensity, 1.0; curve (2), relative intensity, 1.5; curve (3), relative intensity, 2.0; curve (4), relative intensity, 3.0. Dots represent observed points, squares, computed points.

the a.c. photo-potential is a linear function of the intensity of illumination. This is shown in Table I where the photo-potential for each curve is multiplied by the relative intensity and gives a constant for each frequency.

The spectral response curves for the cell for two different frequencies of light flux modulation are shown in Fig. 3. The data for these curves were obtained in the following manner. A Coblentz thermopile (T, Fig. 1) with associated galvanometer was first calibrated by noting the galvanometer deflection when the thermopile was allowed to intercept a known fraction of the radiation from a standard lamp whose power output was known. The photo-e.m.f. of the cell was then measured for each wave-length setting of the spectrometer, after which the thermopile was substituted for the cell and the energy for

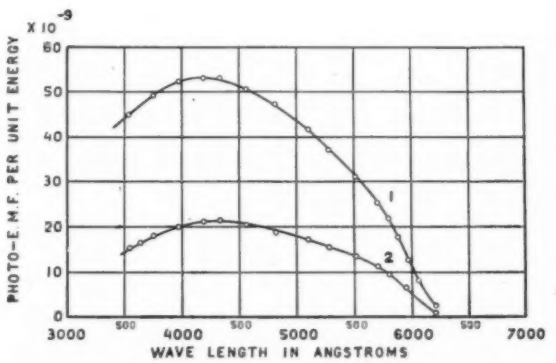

FIG. 3. Spectral response curves of the cell for (1) 1000 cycles, (2) 2000 cycles.

that wave-length determined. From these data, taking into account the absorption of the glass and liquid of the cell, the a.c. e.m.f. per unit energy for each wave-length was computed. It is seen that the frequency of light interruption has no appreciable effect on the wave-length giving maximum sensitivity. The results are in good agreement with those obtained by Barton[4] with steady illumination and d.c. measurements.

The phase shift of photo-potential with respect to the periodically varying illumination was determined by allowing the varying light to fall both on the copper oxide cell and, after reflection by a semi-transparent mirror, on a high vacuum photoelectric cell, and by impressing the two a.c. photo-potentials thus produced (after amplification by two separate amplifiers) on the two pairs of cathode-ray tube plates, and measuring the ellipses which are formed on the screen. Since the time lag of the photoelectric effect in the high vacuum cell is negligibly small,[13] and since the capacity of this cell is of the order of $2\mu\mu f$, the current output of this cell may be considered to be in phase with the light variations (at least for the frequency range of 25 to 10,000 cycles). Consequently, if this current be passed through a non-inductive resistance, the voltage across this resistance may be compared with the terminal a.c. voltage of the copper oxide cell to determine the phase shift of its photo-potential with respect to the varying illumination. If the two voltages are out of phase, an ellipse will be traced on the screen. The sine of the phase angle, B, is given by: $\sin B = \pm ab/A_x A_y$, where a and b are the semi-major and semi-minor axes, respectively, of the ellipse, and A_x and A_y are the amplitudes of the sine waves.[14] The amplifiers were first tested to make sure that no phase shift was introduced by them. The angle of phase shift of the open circuit a.c. photo-potential of the cell obtained in this manner and plotted as a function of the frequency of varying illumination is shown in Fig. 4.

When an inductance was connected across the terminals of the cell and the frequency of light

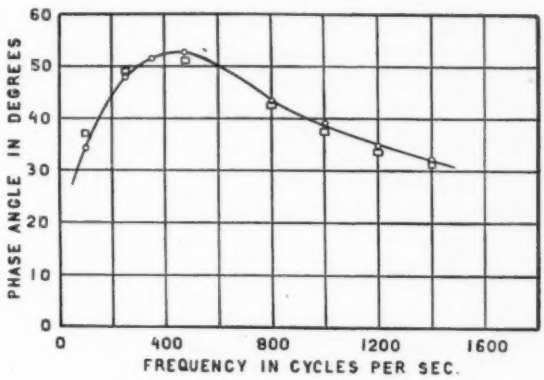

FIG. 4. Phase angle-frequency variation for the cell with sinusoidally varying white light of constant intensity. The points marked by squares are computed.

interruption was varied, a distinct maximum value of the a.c. terminal voltage of the cell was obtained at about 1300 cycles. This indicated a resonant condition in which the cell has considerable capacitance. Measurements with an a.c. bridge supplied with a constant voltage from the oscillator showed this to be true, and also showed that the values of capacity and resistance were not constant with frequency. Each of these quantities was found to vary in accordance with the theory of electrolytic capacity which has been proposed by Warburg.[15] Fig. 5 shows this variation with frequency for the cell. The slope of this straight line is -0.482 which is in close agreement with the value of -0.5 of Warburg. The approximate value of the capacity computed from the previously mentioned resonant condition was $0.18\mu f$, while from Fig. 5 the value at the frequency of 1300 cycles is seen to be $0.2\mu f$. The same light intensity

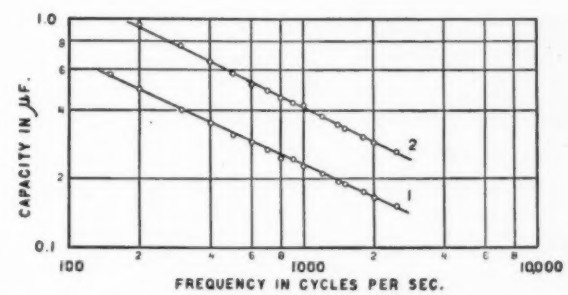

FIG. 5. Capacity-frequency variation for (1) light off, (2) light on.

[13] E. O. Lawrence and J. W. Beams, Phys. Rev. **32**, 478 (1928).

[14] For a detailed description of this method of phase measurements, using the cathode-ray tube see L. A. Wood, R.S.I. **2**, 644 (1931).

[15] E. Warburg, Ann. d. Physik **303**, 493 (1899); **311**, 125 (1901).

used in the resonant condition was used while the cell was connected in the bridge in order to detect any difference in capacity with the light on and off. The light intensity was so small that no difference could be detected in the balance of the bridge, thus indicating that the capacity change was negligible.

However, when one of the electrodes was illuminated by a 100-watt lamp from a distance of 6 inches, a considerable increase in the capacity of the cell was observed. Data for this condition of illumination were obtained and are plotted as line 2 of Fig. 5.

TABLE II. *Parallel capacitances and resistances of electrodes at various frequencies.*

Freq. (cycles/ sec.)	Capacity (μf)	R_p (ohms)	$R_p \cdot f$
100	0.31	1000	100,000
300	0.30	340	102,000
500	0.30	190	95,000
800	0.31	125	100,000
1000	0.30	98	98,000
1600	0.30	64	103,000
2000	0.30	51	102,000

In order to determine the capacity between the copper and the oxide, a plate was made in the usual way and a thin film of gold was deposited on the oxide by evaporation in a high vacuum. The gold thus took the place of the electrolyte, the main difference probably being that it did not penetrate the oxide layer quite as deeply. Three electrodes were prepared in this manner and their equivalent parallel capacitances and resistances were measured with the a.c. bridge. The average values of these quantities are shown in Table II. The capacitance was found to be independent of the frequency while the parallel resistance, R_p, was found to vary inversely with the first power of the frequency, as is shown by the constancy of the product of R_p and f. The equivalent series resistance was found to be negligible and therefore is not shown.

Tests made to determine the thickness of the oxide layer showed that the maximum a.c. photo-potential for a given light intensity occurred with a very definite layer thickness, probably of the order of 10^{-5} cm. It was possible to chemically form a layer so thick that the effect under observation was practically destroyed. With layers 2 mm and more in thickness, made by pressing copper oxide on to the surface of clean copper disks, the photo-effect was so small as to be barely detectable even with maximum amplification. With this thickness it was reasonable to assume that practically no light penetrated to the boundary between the copper and the oxide. This boundary, as well as the method of preparation of the oxide, seems to play a very important rôle in the production of the photo-potential studied in this investigation.

DISCUSSION

If it is assumed, in agreement with Schottky's theory[7] of the dry cuprous oxide cell, (1) that the primary effect of illumination is to cause electrons to pass through the "barrier layer" from the oxide to the copper, and (2) that the number of such electrons is proportional to the intensity of the light, the experimental results can be satisfactorily interpreted if the cell is treated as equivalent to the electrical network shown in Fig. 6. This network corresponds to the cell on open circuit and only the illuminated electrode (a to d in Fig. 6) is represented in detail.

In Fig. 6, C_1 represents the constant electrostatic capacitance between the copper and the

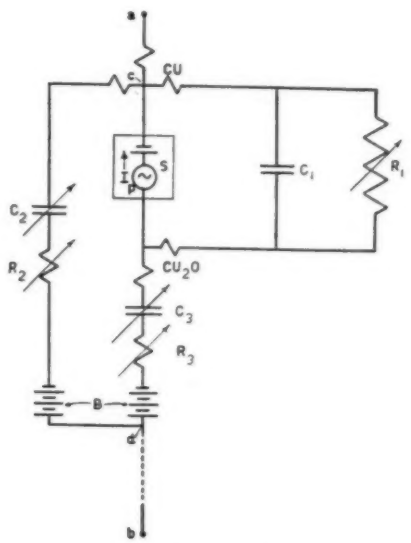

FIG. 6. Electrical network corresponding to the cell on open circuit with sinusoidally varying illumination.

$R_1 = 100{,}000/f$ $R_2 = 13.8 \times 10^3/f^{\frac{1}{2}}$ $R_3 = 10.6 \times 10^3/f^{\frac{1}{2}}$
$C_1 = 0.3\mu f$ $C_2 = 11.5\mu f/f^{\frac{1}{2}}$ $C_3 = 15.0\mu f/f^{\frac{1}{2}}$.

oxide. This capacitance is shunted by a variable resistance R_1 which represents the paths through which the electrons leak back to the oxide. Both of these quantities conform to the values given in Table II. C_2 and R_2 are the electrolytic capacitance and resistance between the copper and the electrolyte, while C_3 and R_3 are the electrolytic capacitance and resistance between the oxide and the electrolyte. The total electrolytic capacitance and resistance for one electrode as measured on the a.c. bridge were divided into the components, C_2, C_3, and R_2, R_3. The values of C_2 and C_3 and of R_2 and R_3 were determined by trial so as to give the best attainable agreement between the results computed for the network and the results actually observed.

To compute the alternating voltage, $\dot{E}_{ad}$ across the illuminated electrode, we put

$$\dot{E}_{ad}=\dot{I}_p\dot{Z}_1\dot{Z}_2/(\dot{Z}_1+\dot{Z}_2),$$

where $\dot{Z}_1$ is the total equivalent impedance from a to d through the right-hand branch and $\dot{Z}_2$ that through the left-hand branch. $\dot{I}_p$ is the a.c. component of current produced by the sinusoidally varying illumination at the surface separating the copper from the cuprous oxide (S in Fig. 6). The symbols and methods of computation are those of the usual a.c. circuit theory in vector algebra form. Assuming a constant value of $\dot{I}_p(=I_p\epsilon^{j0})$, the values of $\dot{E}_{ad}$ for various frequencies were determined from the above relation and the magnitudes and phase angles of this voltage were plotted as squares in Figs. 2 and 4. The value of $\dot{I}_p$ used to determine the points for curve (4) of Fig. 2 was 6.2×10^{-8} amperes while the value of $\dot{I}_p$ for the other curves was less in direct proportion to the relative intensities of illumination used to obtain the experimental points (Fig. 2). The phase angle is the angle between $\dot{E}_{ad}$ and $\dot{I}_p$ and is the angle of the equivalent impedance, $\dot{Z}_1\dot{Z}_2/(\dot{Z}_1+\dot{Z}_2)$.

While the computed and observed results are in excellent agreement throughout the whole range of frequencies it is realized that no rigorous conclusions can be drawn from this fact. Doubtless the results would also fit other hypothetical circuits. But the network chosen is that which seems to correspond most nearly to the actual construction of the cell. It corresponds also to the network proposed and tested by Schottky for the dry copper oxide cell except that resistances and capacitances have been added to correspond to the electrolyte. While these added elements are variable with frequency they are assumed to vary in the way that electrolytic resistances and capacitances are known to behave. The results of this investigation therefore strongly support the view that the source of photoactivity is essentially the same in the wet cuprous oxide cell as in the dry "barrier layer" cell of Grondahl.[16]

In conclusion, the author wishes to express his sincere appreciation and gratitude to Professor Merritt at whose suggestion the problem was undertaken, to Professors Ballard and Bedell, and to many other friends whose suggestions have been extremely helpful.

[16] Direct-current measurements with constant illumination which lead to this same conclusion are briefly reported by R. H. Muller and A. Spector; Phys. Rev. **41**, 371 (1932).

Two Fluid Systems in Porous Media. The Encroachment of Water into an Oil Sand

MORRIS MUSKAT, *Gulf Research and Development Corporation, Pittsburgh, Pa.*

(Received June 6, 1934)

The problem of the encroachment of water into an oil sand is formulated as a new type of potential problem, namely, that of finding potential distributions in two regions of different "constants" ("conductivities") which are separated by a surface, each point of which has a velocity proportional to the vector gradient of the potential at the point, and such that the area swept out by the moving interface assumes the "constant" appropriate to that of the encroaching side of the interface. The cases of strictly linear, radial and spherical systems, in which the shapes of the interfaces are evident from symmetry requirements, are solved in detail and discussed graphically. The zeroth approximation to the general problem which gives the history of a line of fluid particles in a homogeneous system is also treated in detail. Analytical and graphical solutions are presented for (1) systems with elliptical boundaries, (2) an infinite linear source driving fluid into an isolated sink, and (3) the history of a ring of fluid particles travelling from a source to a sink. The relation of the analytical results to the practical problems of the encroachment of water into oil bearing sands is discussed both for the solutions of the general problem and those of the zeroth approximation.

I. INTRODUCTION AND FORMULATION OF THE PROBLEM

THE almost universal occurrence of ground waters in the immediate neighborhood of oil bearing sands lends considerable practical interest to the question of the mutual interaction between the movements of ground waters and the production of the oil from sands contiguous to the water bearing sands. This question also possesses academic interest in that it involves a type of potential problem which, although probably proposed before, does not seem to be treated in the standard texts.

The physical problem may be stated as follows: Given, in a homogeneous sand, a surface S_w, representing a well or fluid outlet surrounded by a body of oil of viscosity μ_1, which is itself bounded externally at a given instant at the surface S_i by water of viscosity μ_2. Over the surface S_e, enclosing the water, a given pressure distribution is applied to "drive" the water and hence the oil into the outlet S_w. The problem is to describe the behavior of S_i and the flow of the oil into S_w.[1] (See Fig. 1.)

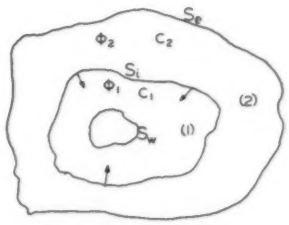

FIG. 1. Diagrammatic representation of a general two-fluid system.

Since we shall be interested only in *viscous* flow through homogeneous sands the basis of the analytical formulation and treatment will be Darcy's law, which states that the velocity of a liquid flowing in a sand is proportional to the pressure gradient acting on the liquid. For three-dimensional systems, where gravity is acting on the liquid, a velocity potential Φ may be introduced as:

$$\Phi = k(p+\gamma g z); \qquad \bar{v} = -c\nabla\Phi, \qquad (1)^2$$

where k is the "permeability" of the sand, p is the pressure, γ the fluid density, g the acceleration of gravity, z the vertical coordinate (directed

[1] Although the problem is essentially one of the transient behavior of S_i we shall neglect throughout all retardation effects upon the flow system due to the change in S_i or in the boundary pressures; the "propagation velocity" of such effects will be taken as infinitely large as compared to the velocity of the fluid particles.

[2] It should be observed that the velocity $v = -c\nabla\Phi$ is equivalent to the volume rate of flow in cc per second as if the fluid were flowing through a medium of 100 percent porosity (fractional void space in the medium), whereas the actual velocity is given by v/f, where f is the true porosity. This particular definition is used here and throughout the remainder of the discussion for analytical convenience, except where specifically noted otherwise.

upwards), $\bar{v}$ the vector velocity and c is the reciprocal of the viscosity of the liquid. Applying the equation of continuity, we have:

$$\nabla \cdot \bar{v} = \nabla^2 \Phi = 0. \tag{1a}$$

It is particularly important to note here that although the permeability defined by Eq. (1), is a constant and gives the dynamical characterization of the *sand* the coefficient c depends on the nature of the liquid. Hence in our physical system, even though the sand is homogeneous throughout, the $c(c_1)$ in the region between S_w and S_i is different from the $c(c_2)$ between S_i and S_e. And furthermore, as S_i sweeps in towards S_w, c_1 is continuously replaced by c_2. Instead of a normal potential problem in which the potentials are to be found on the two sides of a fixed interface separating two regions of different "constants," we have here a potential system with a *moving* interface, the motion of which must be determined simultaneously with the potential distribution. The analogs of an electrostatic system with two dielectrics or heat or current conduction systems composed of parts of different conductivities correspond hydrodynamically to a sand having different values of the permeability on the two sides of S_i but flowing a *single* liquid, so that S_i remains fixed. The inverse case, which is important here, seems to have no analog of practical significance, in other fields of potential theory; it is probably for this reason that this type of potential problem seems to have been given no treatment heretofore.

In giving an analytical formulation of the problem it is convenient to define more closely the motion of the surface S_i. Thus let S_i be represented at the time t by the equation:

$$F(x, y, z, t) = 0. \tag{2}$$

Then at the time $t+\delta t$ the new interface will be given by:

$$F(x+v_x\delta t, y+v_y\delta t, z+v_z\delta t, t+\delta t) = 0 = F(x, y, z, t)$$

where v_x, v_y, v_z are the components of $\bar{v}$ on S_i. We have then at once:

$$\partial F/\partial t + \bar{v} \cdot \nabla F = 0, \tag{3}$$

which is, of course, nothing more than the Kelvin relation governing the motion in a liquid of a surface containing a given set of particles. Applying now Eq. (1) we get:

$$\partial F/\partial t - c\nabla\Phi \cdot \nabla F = 0. \tag{4}$$

The problem may now be formulated as follows: Find the potential functions Φ_1 between a surface S_w and $F(x, y, z, t) = 0$, and Φ_2 between $F(x, y, z, t) = 0$ and a surface S_e such that:

$$\left.\begin{aligned} &\Phi_1 = \Psi_w(x, y, z, t) \text{ on } S_w, \\ &\Phi_2 = \Psi_e(x, y, z, t) \text{ on } S_e, \\ &\left.\begin{aligned} &\Phi_1 = \Phi_2 \\ &c_1 \frac{\partial \Phi_1}{\partial n} = c_2 \frac{\partial \Phi_2}{\partial n} \end{aligned}\right\} \text{ on } F(x, y, z, t) = 0, \\ &\text{where} \\ &\partial F/\partial t - c_{1,2}\nabla\Phi_{1,2} \cdot \nabla F = 0, \end{aligned}\right\} \tag{5}$$

where n is the normal to F and the third and fourth conditions express the requirement of the continuity of the fluid pressures and velocities on F.

It is not difficult to devise a graphical or numerical procedure for solving this problem. Thus by graphical or numerical means Φ_1 and Φ_2 would be computed for the initial position of S_i. Constructing the velocity vectors $-c\nabla\Phi$ over S_i, the position of S_i at a short time interval after the initial instant would be determined. Then with this new position, new values of Φ_1 and Φ_2 would be computed, and so on. Without a numerical integrator, however, this method would probably be quite hopeless.

As to the general problem, even a formal method of solution seems difficult to develop, in view of the fact that even with a stationary S_i the derivation of explicit formal solutions for Φ_1 and Φ_2 are as yet unsolved problems except for very simple geometrical forms for S_e, S_i and S_w. In the following, therefore, we shall treat rigorously only three simple cases in which the symmetry of the problem predetermines the form of F. The discussion of more general cases will be based on the zeroth approximation to the general problem, which is equivalent to taking $c_1 = c_2$

[3] The analogous problems where the normal derivatives on S_w and S_e are specified, or where the boundary conditions as a whole are "mixed" are of minor practical interest. It may also be noted that if gravity be taken into account the continuity of $\Phi/k - \gamma gz$ must be required instead of the continuity of Φ itself. However, if the density difference between the fluids in regions (1) and (2) may be neglected, the condition $\Phi_1 = \Phi_2$ may still be used.

and tracing the path of a line of particles in a *single* fluid in a homogeneous sand.

Before proceeding now with the solution of Eq. (5) for the three cases just mentioned, we may note the following point, which is of considerable practical importance although it applies only to a limited group of problems defined as above. Thus let us suppose that the surfaces S_e and S_w are maintained at uniform potentials Φ_e and Φ_w, respectively. Then since the potential in the region between S_e and S_w is, under these conditions, directly proportional[4] to the difference $\delta\Phi=\Phi_e-\Phi_w$ we may set: $\nabla\Phi=\delta\Phi\nabla_0\Phi$, where $\nabla_0\Phi$ refers to a unit total potential difference. The equation for F may then be written as:

$$\partial F/\partial(t\delta\Phi)-c\nabla_0\Phi\cdot\nabla F=0, \tag{6}$$

and hence since t enters explicitly only in the first term we have that:

$$F=F(x, y, z, t\delta\Phi). \tag{6a}$$

Now the *rate* of advance of the interface S_i, or the rate of "encroachment" of the fluid in region (2) into the region (1), is evidently proportional to $\delta\Phi$. Nevertheless, it is seen from Eq. (6a) that the *shape* of the interface is independent of the rate of encroachment but depends only on the product $t\delta\Phi$. Hence a system with a small $\delta\Phi$ will possess the same family of interfaces F as one where $\delta\Phi$ is large except that they will be crossed at proportionately later instants.

In treating now some special cases of Eq. (5) we shall return, for physical clarity, to the hydrodynamic phraseology and suppose that the region between S_e and $F=0$ is a fluid of viscosity μ_2 "encroaching" on and displacing a fluid of viscosity μ_1 between $F=0$ and S_w, the latter being termed the outflow surface or the "well." On the other hand, we shall not suppose that $\mu_2<\mu_1$ as would be the case if water were encroaching into an oil sand. Rather, we shall discuss the general case where μ_2/μ_1 and hence c_1/c_2 may be both less and greater than 1. The constant c will be called the "conductivity" of the liquid.

[4] It is readily verified that if for a stationary F, Φ_1 and Φ_2 satisfy Eq. (5) with $\Phi_1=\Phi_w$, $\Phi_2=\Phi_e$ on the boundaries, then the solutions Φ_1', Φ_2' for the conditions that $\Phi_1'=\Phi_w'$ on S_w and $\Phi_2'=\Phi_e'$ on S_e are:

$$\Phi_1'=(\Phi_e'-\Phi_w')\Phi_1/(\Phi_e-\Phi_w)+b$$
$$\Phi_2'=(\Phi_e'-\Phi_w')\Phi_2/(\Phi_e-\Phi_w)+b$$
$$b=(\Phi_e\Phi_w'-\Phi_w\Phi_e')/(\Phi_e-\Phi_w).$$

II. Linear Encroachment

The geometrical arrangement is indicated in Fig. 2. Physically, it corresponds to the case where the two fluids are moving in a channel of areal extent that is small as compared to its length. The interface is then a plane element lying normal to the direction of motion.

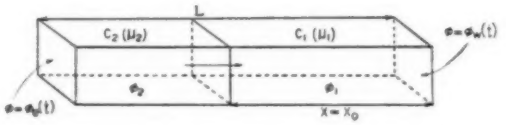

FIG. 2. Diagrammatic representation of a linear two-fluid system.

Turning the problem around as stated in Eq. (5) we may consider first the determination of the surface F, defined for this case by:

$$\frac{\partial F}{\partial t}-c\frac{\partial\Phi}{\partial x}\cdot\frac{\partial F}{\partial x}=0. \tag{7}$$

Because of the simplicity of the problem the potential gradient $\partial\Phi/\partial x$ can be given directly in terms of the position of F. Thus it may be easily verified that at the interface:

$$c_2\frac{\partial\Phi_2}{\partial x}=c_1\frac{\partial\Phi_1}{\partial x}=\frac{c_2c_1(\Phi_e-\Phi_w)}{(c_2-c_1)x+c_1L}. \tag{8}$$

Further, in the present notation:

$$F=x-x_0(t)=0;\quad \frac{\partial F}{\partial t}=-\frac{\partial x_0}{dt};\quad \frac{\partial F}{\partial x}=1.$$

Hence Eq. (7) becomes:

$$\frac{dx_0}{dt}+\frac{c_2c_1(\Phi_e-\Phi_w)}{(c_2-c_1)x_0+c_1L}=0, \tag{9}$$

with the solution:

$$[(c_2-c_1)/2](x_0^2-L^2)+c_1L(x_0-L)$$
$$=-c_1c_2\int_0^t(\Phi_e-\Phi_w)dt, \tag{10}$$

where the initial instant has been taken as that when the interface is at the external boundary $x_0=L$.

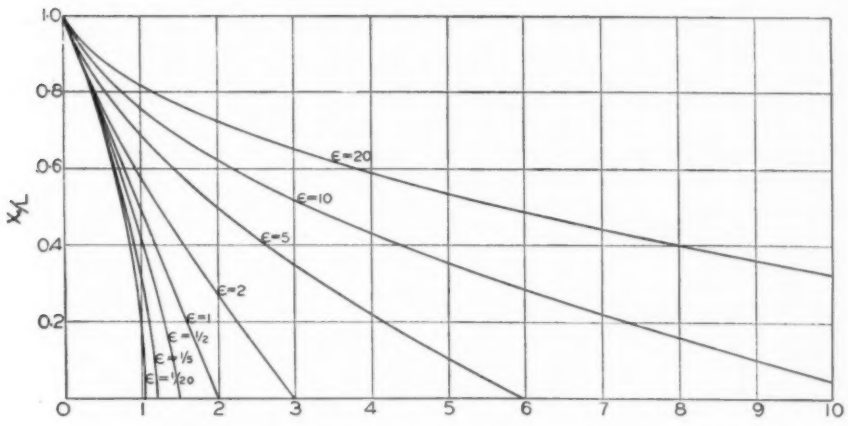

FIG. 3. Progress of the interface in a linear encroachment system as a function of the time. x_0/L = fraction of total length of column occupied at time t (arbitrary units) by the undisplaced liquid. ϵ = (viscosity of encroaching liquid)/(viscosity of liquid displaced).

For the case when the difference in the boundary values, $\Phi_e - \Phi_w$, is kept constant Eq. (10) takes the form:

$$(x_0/L-1)\{(x_0/L)(1-\epsilon)+\epsilon+1\} + (2\Delta\Phi/L^2)c_1 t = 0; \qquad \epsilon = c_1/c_2. \quad (11)$$

The fact that t enters in the combined term $2c_1\Delta\Phi t/L^2$ shows that for corresponding states of the encroachment the time required to reach these states varies inversely as the average potential gradient $\Delta\Phi/L$, directly as the total length of the system L, and directly as the viscosity of the displaced fluid. With this in mind we may study the details of the encroachment by simply plotting x_0/L against t, setting, for convenience, $2c_1\Delta\Phi/L^2$ equal to unity. The curves so computed are given in Fig. 3 for various values of ϵ. The individual curves are parabolic, as Eq. (11) indicates. And as is to be expected, the encroachment of the interface is rapidly accelerated with decreasing x_0 if the encroaching fluid has the higher conductivity, ($\epsilon<1$) whereas it is retarded if the encroaching fluid has the lower conductivity ($\epsilon>1$). The actual flux through the system as the interface moves toward $x_0=0$ is given essentially by the slopes of the curves of Fig. 3 and are presented more explicitly in Fig. 4. They have been computed from the equation:

$$dy/dt = -1/2(\epsilon+y-\epsilon y): \qquad y = x_0/L. \quad (9a)$$

The marked acceleration of the interface and corresponding increase in flux through the system for the cases $\epsilon<1$ is clearly shown in Fig. 4 in the rapid rise of $-dy/dt$ as y approaches 0.

The total time required for the interface to move across the system is given by the value of t at $x_0=0$ or:

$$t_{(\max)} = \frac{(1+c_1/c_2)L^2}{2\Delta\Phi c_1}, \quad (11a)$$

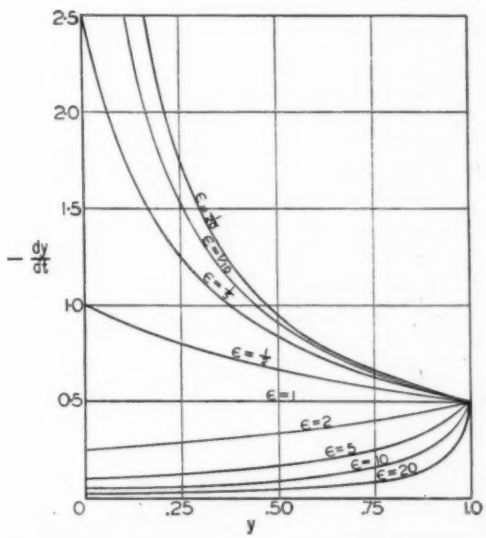

FIG. 4. Rate of fluid displacement ($-dy/dt$) in a linear encroachment system. y = fraction of total length of column occupied by undisplaced liquid. ϵ = (viscosity of encroaching liquid)/(viscosity of liquid displaced).

which is evidently less or larger than the single fluid "encroachment time"

$$t^{\circ}_{(\max)} = L^2/c_1\Delta\Phi \text{ as } c_1/c_2 < 1 \text{ or } c_1/c_2 > 1. \quad (11b)$$

It may be noted, too, from Eq. (11a) that the maximum decrease in $t_{(\max)}$ due to the encroachment is attained when $c_2 \gg c_1$ so that

$$t_{(\max)}/t^{\circ}_{(\max)} \sim \tfrac{1}{2}. \quad (12)$$

Hence in this limiting case the drive by the encroaching fluid would cut by half the time required for the column of original liquid to pass through the system.

III. Two-Dimensional Radial Encroachment

As indicated in Fig. 5 we assume for this case complete radial symmetry. Thus the functions

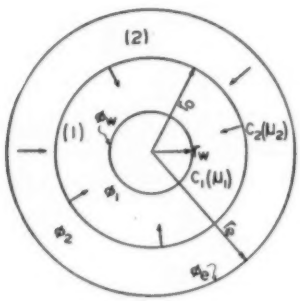

FIG. 5. Diagrammatic representation of a radial two-fluid system.

Φ_2 and Φ_1 are uniform over the external and internal boundaries, defined by the cylinders $r = r_e$, $r = r_w$, respectively.

We may set then:

$$F = r - r_0(t) = 0; \quad \nabla\Phi \cdot \nabla F = \frac{\partial\Phi}{\partial r}\frac{\partial F}{\partial r}. \quad (13)$$

Hence the equation for F becomes:

$$\partial r_0/\partial t + c(\partial\Phi/\partial r)_{r_0} = 0. \quad (14)$$

Now from elementary potential theory it may be shown that if the interface between the two media is at r_0, then:

$$c_1\frac{\partial\Phi_1}{\partial r} = \frac{c_1(\Phi_e - \Phi_w)/r_0}{\log(r_0/r_w) + \epsilon\log(r_e/r_0)} = c_2\frac{\partial\Phi_2}{\partial r}; \quad \epsilon = c_1/c_2. \quad (15)$$

Putting this in Eq. (14) and integrating:

$$r_0^2\{b - a + a\log r_0^2\} = -4c_1\int_0^t \Delta\Phi dt + \text{const.} \quad (16)$$

where:

$$a = 1 - \epsilon; \qquad b = \log r_e^2/r_w^2 - a\log r_e^2.$$

If, now, it be assumed that $\Delta\Phi$ is constant and that at $t = 0$, $r_0 = r_e$ Eq. (16) may be rewritten as:

$$\frac{r_0^2}{r_e^2}\left\{\log\frac{r_e^2}{r_w^2} - a + a\log\frac{r_0^2}{r_e^2}\right\} = -4c_1 t\frac{\Delta\Phi}{r_e^2} + \log\frac{r_e^2}{r_w^2} - a. \quad (17)$$

Setting $4c_1\Delta\Phi/r_e^2 = 1$ and adding the numerical assumption that $r_e/r_w = 2000$, which is of the order of magnitude involved in cases of practical interest, curves representing Eq. (17) for various values of ϵ are plotted in Fig. 6. It will be observed that as the encroachment proceeds, i.e., as r_0/r_e decreases, the rate of encroachment is accelerated. It should be noted, however, that here the actual acceleration is the resultant of the natural acceleration due to the converging character of the flow and the effect of replacing the original fluid of conductivity c_1 by that of conductivity c_2. Thus, for $\epsilon = c_1/c_2 < 1$ the second effect continually augments the natural acceleration given by the case $\epsilon = 1$. On the other hand, for $\epsilon > 1$ the encroaching medium opposes the natural acceleration and the slopes are diminished. And for ϵ so large that $\epsilon - 1 \geq \log(r_e/r_w)$ this second opposing effect may even counterbalance the natural acceleration and lead to a temporary decrease in the rate of encroachment. This is already shown in the present case in the curve for $\epsilon = 10$ and is still more prominent when $\epsilon = 20$, where a marked inflexion may be noted at about $r_0/r_e = 0.55$.

The rate of fluid flow during the encroachment, which is proportional to $\partial r_0^2/\partial t$, may be computed by the equation:

$$\frac{\partial(r_0/r_e)^2}{\partial t} = \frac{-2c_1\Delta\Phi/r_e^2}{(1-\epsilon)\log(r_0/r_e) + \log(r_e/r_w)}. \quad (18)$$

Setting again: $4c_1\Delta\Phi/r_e^2 = 1$, and $r_e/r_w = 2000$, Eq. (18) is plotted in Fig. 7 for various values

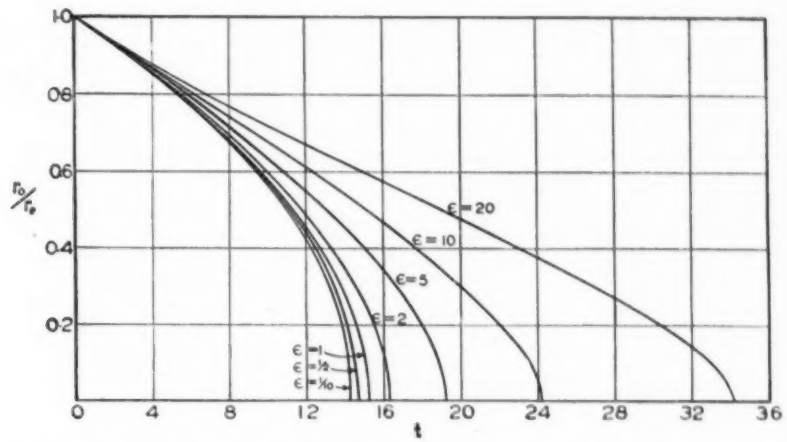

FIG. 6. Progress of the interface in a radial encroachment system as a function of the time. r_0 = radius of interface at time t. r_e = radius of external boundary. ϵ = (viscosity of encroaching liquid)/(viscosity of liquid displaced). r_w = well radius = $r_e/2000$.

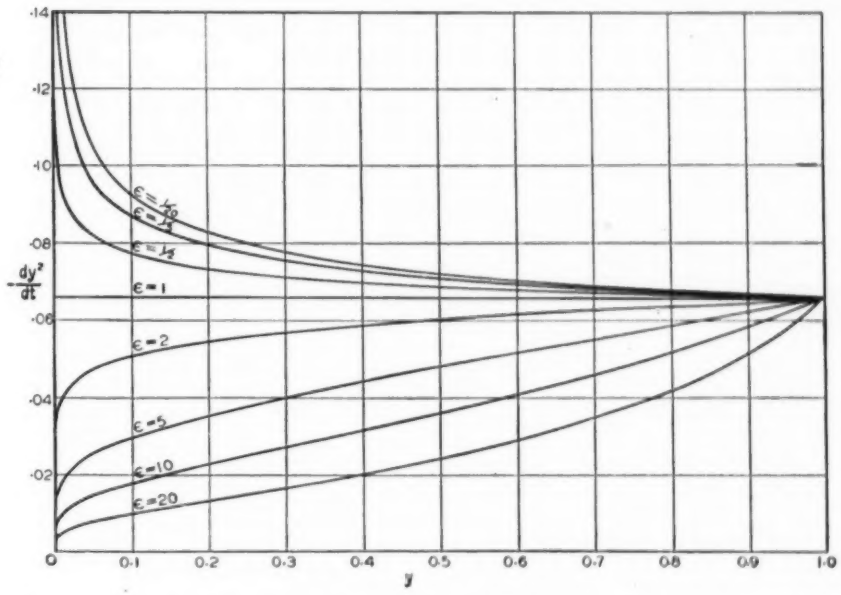

FIG. 7. Rate of fluid displacement ($-dy^2/dt$) in a radial encroachment system. y = (radius of interface)/(radius of external boundary) ϵ, r_w same as in Fig. 6.

of ϵ, with r_0/r_e replaced by y. The interpretation is similar to that of Fig. 4.

The total time required for the interface to move from $r_0 = r_e$ to $r_0 = r_w$ is by Eq. (17):

$$t_{(\max)} = \frac{r_e^2}{4c_1\Delta\Phi}\left[\left(1-\epsilon\frac{r_w^2}{r_e^2}\right)\log\frac{r_e^2}{r_w^2}+(\epsilon-1)\left(1-\frac{r_w^2}{r_e^2}\right)\right] \leq \frac{r_e^2}{4c_1\Delta\Phi}\left[\log\frac{r_e^2}{r_w^2}+\epsilon-1\right] : \frac{r_w}{r_e} \ll 1. \qquad (19)$$

It appears, therefore, that for $\epsilon < 1$ the decrease in $t_{(\max)}$ due to the entry of the higher conductivity fluid cannot exceed the value $1/2 \log (r_e/r_w) = 0.066$ for $r_e/r_w = 2000$, whereas in the corresponding linear flow problem the encroachment could cut $t_{(\max)}$ into ½ of that

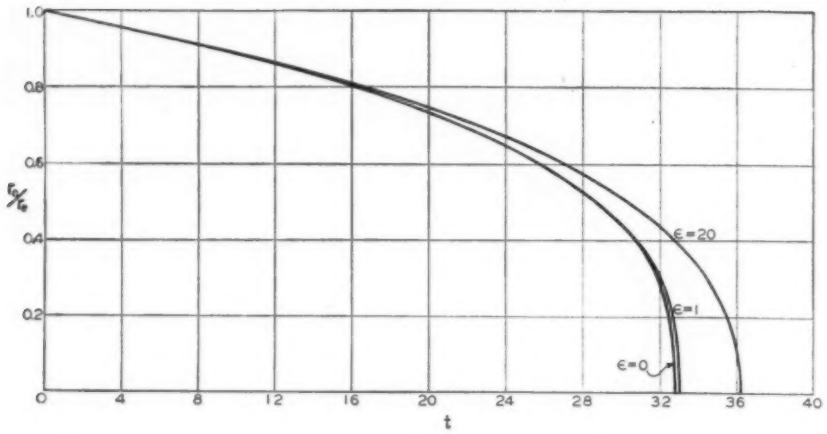

FIG. 8. Progress of the interface in a spherical encroachment system. Notation same as in Fig. 6.

without encroachment. Of course, as Fig. 7 shows, after the encroachment is completed the resultant flux is inversely proportional to ϵ as it would be in a homogeneous system.

IV. SPHERICAL ENCROACHMENT[5]

Since the analysis for this case follows exactly the lines of that for the two-dimensional radial encroachment, only the results will be given. They are, with the notation used in Section III:

$$\frac{c_1\Delta\Phi t}{r_e^2}=\left(\frac{r_0}{r_e}\right)^2\left\{\frac{r_0}{3r_e}\left(\epsilon-\frac{r_e}{r_w}\right)+\frac{1-\epsilon}{2}\right\}$$

$$-\frac{1-\epsilon}{2}+\frac{1}{3}\left(\frac{r_e}{r_w}-\epsilon\right) \quad (20)$$

and

$$\frac{\partial(r_0/r_e)^3}{\partial t}=\frac{3c_1\Delta\Phi/r_e^2}{(1-\epsilon)/(r_0/r_e)+\epsilon-r_e/r_w}. \quad (21)$$

Setting: $c_1\Delta\Phi/r_e^2=1$; $r_0/r_e=y$; $r_e/r_w=100$, Eq. (20) is plotted in Fig. 8. The difference between this figure and Figs. 6 and 7 for radial encroachment may be readily explained on noting that here the natural convergence of the flow is much more pronounced than in the case of the radial flow.

The time of passage of the interface from $r_0=r_e$ to $r_0=r_w$ is easily seen from Eq. (20) to be given by:

$$t_{(\max)}\cong(\epsilon-1)/2+\tfrac{1}{3}(r_e/r_w-\epsilon), \quad (22)$$

unless $r_w/r_e=0(1)$.

The ratio of this to the $t_{(\max)}$ in a homogeneous system ($\epsilon=1$) is then:

$$\frac{t_{(\max)}}{t_{(\max)}(\epsilon=1)}=\frac{(\epsilon-1)/2+\tfrac{1}{3}(r_e/r_w-\epsilon)}{\tfrac{1}{3}(r_e/r_w-1)}$$

$$\cong 1-\frac{3}{2}\frac{r_w}{r_e}(1-\epsilon) \quad (23)$$

for $\epsilon<1$. The effect, here, is therefore, still smaller than in the case of the radial flow. Thus for $r_w/r_e=0.01$, $t_{(\max)}$ will be reduced by no more than 1.5 percent over that in a homogeneous system, even if $\epsilon=0$. From a qualitative point of view this small effect is to be expected from the fact that the pressure gradients are very concentrated for small value of y and the higher conductivity fluid reaches the small values of y only at the very last stages of the encroachment which consume only a negligible fraction of the total time required for the whole process to be completed.

Finally it may be observed that although the solutions of these problems just treated have been formally based upon the general differential equation for F they can also be solved directly by simply requiring that the rate of volume swept by the interface is the instantaneous rate of outflow through the systems. For these cases this direct requirement is fully equivalent to the

[5] Since this is essentially a three-dimensional problem we must either neglect the effect of gravity entirely or suppose that the two fluids are of the same density.

differential equation for F. However, when the interface is not of such simple forms as cylindrical or spherical surfaces the general formulation would be necessary.

V. The History of a Line of Fluid Particles in a Homogeneous System

The problems treated above have all possessed such a degree of symmetry that the forms of the interface have been immediately evident. In less simple problems, the form of the interface as well as its instantaneous position must be determined simultaneously. Unfortunately the practical difficulties of solving such general cases, defined by Eq. (5) have been too great to permit the formulation of a general procedure. The discussion of cases of more complex geometrical forms will, therefore, be based on the zeroth approximation to the general problem in which it is assumed that $c_1 = c_2$ and the history of a line of particles is traced for a single fluid in a homogeneous sand. Although from an analytical point of view this zeroth approximation or the corresponding problem in a completely homogeneous system has in itself no relation to that of the moving interface in a non-homogeneous system, nevertheless it is physically evident that at least a qualitative idea of the nature of the solution should be afforded by the simplified problem. We shall, therefore, present a discussion of the problem of the motion of a line of fluid particles in a homogeneous system, that is, one in which the conductivity, c, is everywhere a constant.

By Eq. (5) the function defining the line of particles at any time t will satisfy the equation:

$$\partial F/\partial t - \nabla\Phi\cdot\nabla F = 0, \tag{24}$$

where c has been taken as unity. Although there are available formal methods for treating any first-order linear partial differential equation we shall proceed here with a method which makes direct use of the fact that the multiplier of ΔF is the gradient of a potential function.

Since the system is homogeneous it will be steady if the boundary conditions are independent of the time. There will, therefore, be in the system a permanent set of equipotential and stream line surfaces. If then, an orthogonal set of curvilinear coordinates are introduced and defined by u, v, w (x, y, z), $\nabla\Phi\cdot\nabla F$ will take the form:

$$\nabla\Phi\cdot\nabla F = \frac{\partial\Phi}{\partial u}\cdot\frac{\partial F}{\partial u}|\nabla u|^2 + \frac{\partial\Phi}{\partial v}\cdot\frac{\partial F}{\partial v}|\nabla v|^2 + \frac{\partial\Phi}{\partial w}\cdot\frac{\partial F}{\partial w}|\nabla w|^2. \tag{25}$$

Hence if one of the set of surfaces, u, for example, is taken as the set of equipotential surfaces, $\Phi = \text{const.}$ the last two terms in $\nabla\Phi\cdot\nabla F$ will vanish and the equation for F will take the form:

$$\partial F/\partial t - |\nabla u|^2 \partial F/\partial u = 0, \tag{26}$$

where $|\nabla u|^2$ is to be expressed in terms of u, v, w.

Eq. (29) is easily integrated with the result

$$F = \alpha t + \alpha \int du/|\nabla u|^2 + g(v, w) = \text{const.}, \tag{27}$$

where g is an arbitrary function to be adjusted so that F assumes its initial form at $t = 0$ (α may be always taken as unity). All the latter history of the particles lying initially on this line may then be found by simply plotting Eq. (27) for successive values of t.

This theory will now be illustrated by several examples. For simplicity we shall confine ourselves to two-dimensional problems. For such cases the curvilinear coordinates may be taken as the equipotential curves $\Phi = \text{const.}$ and stream lines $\Psi = \text{const.}$ Eq. (27) then may be written as:

$$F(\Phi, \Psi) = t + \int d\Phi/|\nabla\Phi|^2 + g(\Psi) = \text{const.} \tag{28}$$

(1) Elliptical boundaries (see Fig. 9)

If the boundaries of the system are confocal ellipses the equipotentials and stream line functions may be defined by the well-known relations:

$$\left.\begin{aligned} x + iy &= c\cosh(\Phi + i\Psi), \\ x &= c\cosh\Phi\cos\Psi, \\ y &= c\sinh\Phi\sin\Psi. \end{aligned}\right\} \tag{29}$$

From these equations it may be readily verified that:

$$|\nabla\Phi|^2 = 2/c^2(\cosh 2\Phi - \cos 2\Psi). \tag{30}$$

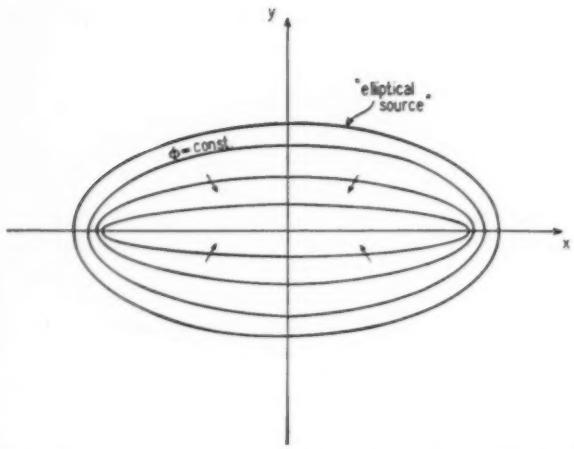

FIG. 9. Diagrammatic representation of an elliptical encroachment system.

Hence by Eq. (28):

$$F=t+(c^2/2)(\tfrac{1}{2}\sinh 2\Phi-\Phi\cos 2\Psi)+g(\Psi)=\text{const.} \quad (31)$$

If we suppose that at

$$t=0 : \Phi=\Phi_0(\Psi)$$

we get:

$$t=(c^2/2)[\tfrac{1}{2}(\sinh 2\Phi_0-\sinh 2\Phi)+(\Phi-\Phi_0)\cos 2\Psi]. \quad (32)$$

For the case when the line of particles lies initially upon the ellipse with the major and minor axes bearing the ratio 2 : 1, i.e., for $\Phi_0(\Psi)=0.5493$ and for $c=1$ the curves $F(\Psi, \Phi, t)$ are plotted in Fig. 10. The equipotentials of the system have been drawn in Fig. 10 as the dotted curves $\Phi=$const. Here we see at once that the surfaces $t=$const. do not follow the equipotential surfaces. In fact, it may be proved quite generally that, except under special conditions which happen to be satisfied by the simple examples treated above (linear, radial and spherical flow), the surfaces $t=$const. will not follow the equipotentials even if $t=0$ coincides with an equipotential surface. The paths which the fluid particles, initially on $\Phi=0.5493$, take in reaching the "well," $\Phi=0$, are indicated by the stream lines $\Psi=$const.

Although we shall not enter into a detailed discussion of the curves $t=$const., they may be readily interpreted on noting that the square of the fluid velocity at any point in the system is given by Eq. (30) and that the slopes of the curves $t=$const. in the (Φ, Ψ) coordinate system are given by:

$$\frac{dy}{dx}=-\frac{\cosh 2\Phi-\cos 2\Psi+4(\Phi_0-\Phi)\operatorname{ctnh}\Phi\sin^2\Psi}{\tan\Psi\operatorname{ctnh}\Phi(\cosh 2\Phi-\cos 2\Psi)-2(\Phi_0-\Phi)\sin 2\Psi}. \quad (33)$$

(2) Line drive into a well

A case which shows in a simpler manner some of the interesting features of the general problem of encroachment is that in which the history of a line of particles is traced as it is brought into a well at a given distance from a line source.

The equipotential and stream line distributions for this case may be conveniently derived from the complex variable relation:

$$\Phi+i\Psi=q\log\frac{z-ih}{z+ih}+\varphi_0, \quad (34)$$

so that

$$\left.\begin{aligned}\Phi&=\frac{q}{2}\log\frac{x^2+(y-h)^2}{x^2+(y+h)^2}+\varphi_0\\ \Psi&=q\tan^{-1}\frac{-2hx}{x^2+y^2-h^2},\end{aligned}\right\} \quad (35)$$

where q is a constant determining the flux and φ_0 is the potential over $y=0$. Introducing the notation

$$(\varphi_0-\Phi/q)=\eta;\ \ \Psi/q=\xi;\ \ x/h=x;\ \ y/h=y, \quad (36)$$

it follows from Eq. (35) that

$$y=\sinh\eta/(\cosh\eta\pm\cos\xi);$$

$$x=\pm\sin\xi/(\cosh\eta\pm\cos\xi). \quad (37)$$

Since the streamlines near the y axis are to be chosen so that $x>0$ for $\Psi>0$ the upper sign must be used. $|1/\nabla\Phi|^2$ may then be expressed as:

$$1/|\nabla\Phi|^2=h^2/q^2(\cosh\eta+\cos\xi)^2. \quad (38)$$

Hence by Eq. (28), the equation of the line of particles at time t is given by:

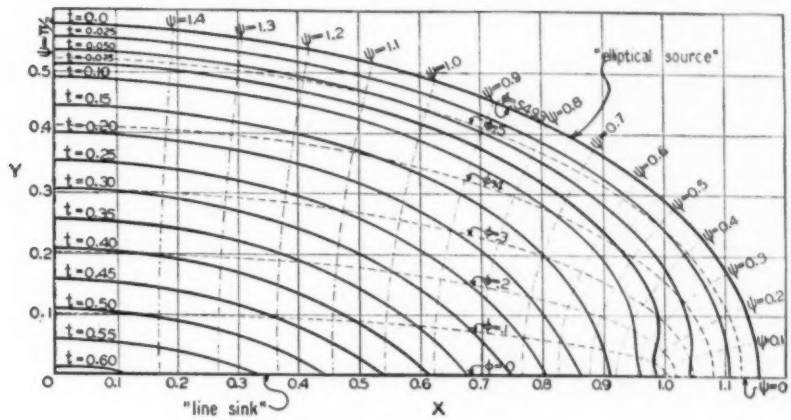

FIG. 10. History of a line of fluid particles as they emerge from an elliptical source and travel in a homogeneous medium toward a confocal line sink. Φ = const. equipotentials; Ψ = const. stream lines.

$$F(\xi, \eta, t) = t - \frac{h^2}{q \sin^2 \xi}\left\{\frac{\sinh \eta}{\cos \xi + \cosh \eta} - 2 \operatorname{ctn} \xi \tan^{-1} \tan \frac{\xi}{2} \tanh \frac{\eta}{2}\right\} + g(\xi) = 0. \tag{39}$$

If the initial position of the line of particles of interest is the x axis, i.e., the curve $y = 0$, we must have $g(\xi) = 0$. The final result is then:

$$t = \frac{h^2}{q \sin^2 \xi}\left[\frac{\sinh \eta}{\cos \xi + \cosh \eta} - 2 \operatorname{ctn} \xi \tan^{-1} \tan \frac{\xi}{2} \tanh \frac{\eta}{2}\right]. \tag{40}$$

Along the y axis, i.e., $\xi = 0$, Eq. (40) is indeterminate. The usual methods give:

$$t(x = 0) = (yh^2/6q)(3 - y^2). \tag{41}$$

And when ξ is small an expansion of t in powers of ξ may be used in the numerical treatment.

Setting $h = q^2 = 1$ the curves t = const. are plotted in Fig. 11 in a Cartesian coordinate system. At the initial instant the line of particles whose history has been traced, lie on the x axis. For convenience the potential of the x axis has been taken as $\Phi = 5.0$. (This corresponds to setting $\varphi_0 = 5.0$.) The other equipotentials, Φ = const., and the stream lines Ψ = const., defined by Eqs. (35), have been drawn as the dotted curves. As may be readily verified, these are a system of mutually orthogonal circles, their centers lying upon the y and x axes, respectively.

As might have been anticipated, the curves t = const. ≤ 0.33 have maximal coordinates along the y axis. This simply means that the particles near the y axis move faster than those distant from it; and this, of course, is a consequence of the higher gradients of the potential along the y axis. Furthermore, as the particles approach the well, at $y = 1$, they move into regions of increasing gradients and their velocities are continually accelerated. The maximum for increasing values of t, becomes more and more pronounced until the limiting case $t = 0.33$, when the particles along the y axis experience such large gradients that the curve is drawn in as a cusp as the first particle, of those initially on the line $y = 0$, enters the well.

Once the curve t = const. goes through the point of convergence, $y = 1$, all the succeeding curves will also pass through this point, since all the streamlines must end there. On the other hand, since this point remains fixed and the other points along the curves t = const. must follow the stream lines, the curves for $t > 0.33$ will flatten out again near the y axis and those for high values of t will even follow the streamlines above the point (0, 1) and produce a cusp in the direction opposite to that for $t = 0.33$.

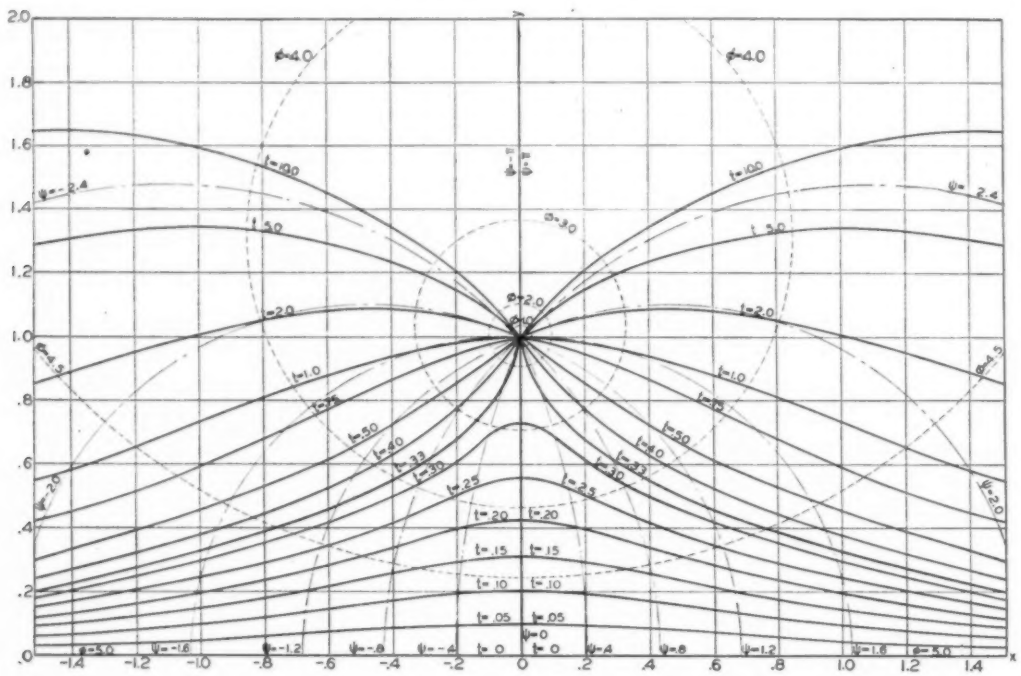

FIG. 11. History of a line of fluid particles as they emerge from an infinite line source and travel in a homogeneous medium toward an isolated sink (well). Φ = const. equipotentials; Ψ = const. stream lines.

These details may be verified by reference to Fig. 11.

Finally it may be of interest to note the order of magnitudes involved when the above analysis is applied to a system with dimensions of practical interest. Thus from Eq. (35) it is clear that if the line drive is located at a distance h from the well and the total pressure differential is Δp the times involved will be:

$$\bar{t}=\frac{fh^2\mu t}{k\Delta p}\sinh^{-1}\frac{h}{r_w}, \tag{42}$$

where t is the time used in Fig. 12, r_w is the well radius, k is the actual permeability of the sand, f is its porosity and μ the viscosity of the fluid. Hence if

$$\left.\begin{aligned} &k=1 \text{ darcy};^{6} \quad f=0.20; \quad \mu=1 \text{ centipoise};\\ &h=500'; \quad r_w=1/4'; \quad \Delta p=700 \text{ lb.}\end{aligned}\right\} \tag{43}$$

the time $t=\frac{1}{3}$ for the line of particles to first reach the well will correspond to:

$$\bar{t}=31.2 \text{ days.} \tag{44}$$

[6] 1 darcy is a permeability of 1 cc/sec./cm²/atmos./cm for a fluid of 1 centipoise viscosity (approx. water at 20°C).

(3) Direct drive between two wells

As a final example we shall treat the problem of the direct flooding between two wells, alone in an infinite two-dimensional reservoir. That is, we shall trace the history of a line of particles just emerging from a "driving" or "flooding" well at $(0, -h)$ as they travel towards the output well at $(0, h)$. As in the previous two cases, the instantaneous curves assumed by these particles will correspond in a real flooding problem to the water-oil interfaces at the corresponding instants, except, of course, for the correction due to the viscosity difference between the oil and water which shall be neglected here.

Upon a little inspection it will be observed that the potential distribution for the present problem will be identical with that for the case of the line drive into a well and that the peculiar feature characterizing the present problem will then enter only in the statement of the initial conditions, i.e., that instead of requiring that at $t=0$ the line of particles lie on the x axis, we shall require that at $t=0$ they have just emerged from the "flooding" well at $(0, -h)$. Specifically we require that:

$$\eta=\eta_0 \quad \text{at} \quad t=0, \tag{45}$$

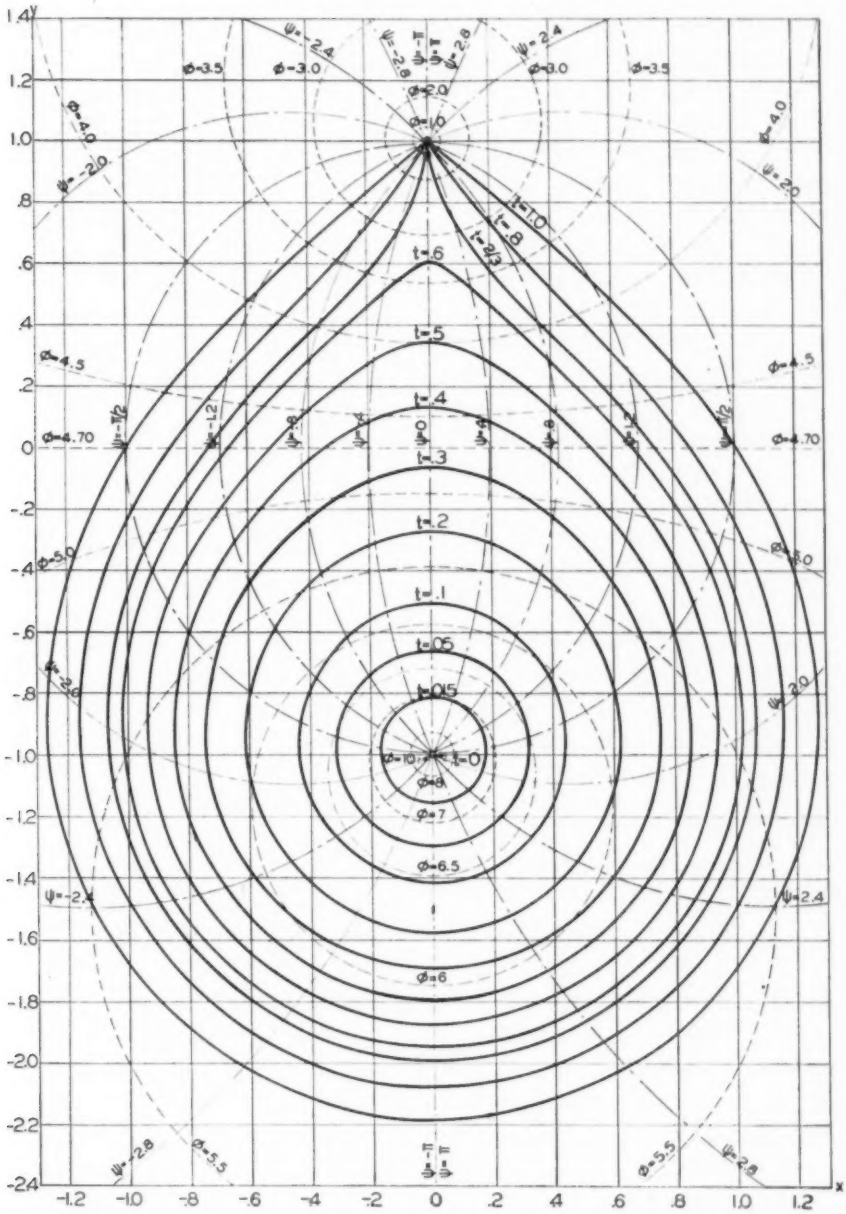

FIG. 12. History of a ring of fluid particles as they emerge from a source and travel towards a sink. Φ = const. equipotentials; Ψ = const. stream lines.

where η_0 is an equipotential encircling the input well or even coincident with it. With this requirement, $g(\xi)$ in Eq. (39) may be determined, and we have finally:

$$t=\frac{h^2}{q\sin^2\xi}\left[\frac{\sinh\eta}{\cosh\eta+\cos\xi}-2\operatorname{ctn}\xi\tan^{-1}\tan\frac{\xi}{2}\tanh\frac{\eta}{2}\right]$$

$$-\frac{h^2}{q\sin^2\xi}\left[\frac{\sinh\eta_0}{\cosh\eta_0+\cos\xi}-2\operatorname{ctn}\xi\tan^{-1}\tan\frac{\xi}{2}\tanh\frac{\eta_0}{2}\right]. \quad (46)$$

Again choosing $h^2=q=1$ and taking the initial position of the line of particles as an equipotential circle of radius 1/200 of the distance between the wells, which implies that $\eta_0=-\sinh^{-1}100$, the curves of Fig. 12 were computed and plotted.

Perhaps the most striking feature of these curves is their small distortion from complete radial symmetry. Although the curves are actually oval-shaped the sharpness of the advancing front of the curves does not become appreciable until the particle on the y axis—the line of centers between the walls—has travelled to the midpoint between the wells. This behavior may, however, be readily explained as in the case of the "line drive," by taking into account the details of the potential distribution in the system, as shown by the curves $\Phi=\text{const.}$

The order of magnitude represented by the values of $\bar{t}\leq\frac{2}{3}$ which correspond in Fig. 12 to the total time required for the fluid from the input well to first reach the output well, is given by:

$$\bar{t}=\frac{2fh^2\mu t}{k\Delta p}\sinh^{-1}\frac{h}{r_w}, \tag{47}$$

where the notation is that used in Eq. (42).

Hence with the same constants as in Eq. (43) except that here $2h=500'$: $\bar{t}=42.9t$ days, so that $t=\frac{2}{3}$ gives: $\bar{t}=28.6$ days for the line of particles to first reach the output well. Comparing this result with Eq. (44) we see that it would take less time for fluid to pass between two wells than between a line drive and a well separated by the same distance as in the two well system, and flowing under the same pressure differential.

Although the above illustrations show that, in principle, the problem of the homogeneous fluid system can be solved whenever the equipotentials and stream lines can be found explicitly, the analytical integration of Eq. (27) may involve a very difficult and laborious procedure for systems more complex than those chosen for the illustrations. A graphical integration, however, can always be carried out once the stream lines have been found and $|\nabla\Phi|^2$, or the square of the velocity, along them can be computed. On the other hand, even this work may be avoided in getting the history of the line of particles by using the ingenious and extremely simple electrolytic model devised by Wyckoff and Botset.[7]

VI. Area of Encroachment

Although the analytical prediction of the detailed shape of the encroaching line of particles, as given in Figs. 10, 11 and 12, necessitates a complete tracing of the equipotentials and stream lines in the system, the total area swept out at any time and, in particular, at the time the first particle reaches the output wells, may be obtained quite simply even in relatively complex systems. For this area must evidently equal the rate of influx of fluid from the source of the system multiplied by the time in question. Thus for the total area swept out by the time the fluid particles first reach the outlet sink we have for the cases treated above:

(1) Elliptical boundaries

$$t=t(\Phi=0,\ \Psi=0)=(fc^2/2)(\tfrac{1}{2}\sinh 2\Phi_0-\Phi_0), \tag{48}$$

by Eq. (32). As the rate of influx into the quadrant of Fig. 10 is $\pi/2$, the total area swept out by the time t is

$$A=\pi t/2f=(\pi c^2/4)(\tfrac{1}{2}\sinh 2\Phi_0-\Phi_0). \tag{48a}$$

The fraction of the total area of the flooding system is evidently A divided by A_0, the area of the quadrant, i.e., by $A_0=\pi c^2\sinh 2\Phi_0/8$. Hence

$$A/A_0=1-2\Phi_0/\sinh 2\Phi_0=0.176 \tag{49}$$

with $\Phi_0=0.5493$, as assumed in Fig. 10.

(2) Line drive into a well

Here, by Eq. (41) the time for the particles to first reach the output well is $t=h^2f/3q$. But, by Eq. (35) the flux rate in the system is $2\pi q$, so that the total area swept out is:

$$A=2\pi h^2/3. \tag{50}$$

(3) Direct drive between two wells

Similarly, for the direct drive between two wells it is found that

$$A=4\pi h^2/3. \tag{51}$$

[7] Cf. R. D. Wyckoff and H. G. Botset, Physics 5, 265 (1934).

For more complex systems, as two-dimensional flooding networks, consisting of arrays of wells, the essential problem may again be reduced to that of finding the time t for the particles to first reach the output well. From the symmetry of the system it is possible in most cases of practical interest to find directly the stream line along which these fastest particles will travel. The potential distribution, on the other hand, may be written in the form:

$$\Phi = (q/4\pi) F(x, y) \tag{52}$$

where q represents the flux rate from a representative input well. t, then, is given by

$$t = (4\pi f/q) \int_a^b ds/(\partial F/\partial s), \tag{52a}$$

where ds is an element of the stream line of shortest time and a and b represent adjacent input and output wells. The total area swept out per input well by the time the flooding liquid first reaches the output well, is then:

$$A = qt/f = 4\pi \int_a^b ds/(\partial F/\partial s). \tag{53}$$

Since the potential gradient $\partial F/\partial s$ need be computed or known only along the particular stream line of shortest time, A can be found rather easily even though the complete tracing of the flood may be extremely laborious. The detailed analysis of this type has been given elsewhere for various types of flooding networks.[8]

VII. Practical Applications

In order to bring out more clearly the practical significance of the above analysis we shall briefly reinterpret and summarize them from a more practical point of view. We shall, therefore, consider in particular the case where $c_2/c_1 > 1$, corresponding to water encroaching in an oil sand.

Thus, first, it is to be noted from Figs. 4 and 7 that although the production rate will increase as the encroachment progresses, the increase in a nonlinear system will not become appreciable until the encroaching fluid approaches quite closely to the outlet well. In fact, for a radial system in which $c_2/c_1 = 20$, the production rate will not be doubled until 99.97 percent of the oil will have been displaced by the encroaching water, so that a real observable rise in the production rate will imply that the water is in the immediate vicinity of the well. Furthermore, as the geometrical convergence of the system increases, the nature of the encroaching fluid becomes of decreasing importance, since it will come into play only in the very last stages of the encroachment which consume but a small fraction of the total time required for the whole process to be completed. Thus it would take 118.5 days for a linear column of 500 feet to be swept out in a unit permeability sand by water (1 centipoise viscosity) encroaching at 700 lb. pressure differential if the viscosity of the oil is 20 times as great, but as high as 225.8 days if the water had the same viscosity as the oil. On the other hand, for the analogous *radial* system—a 500 foot disk of sand about a $\frac{1}{4}$ foot well—it would take 804.5 days for the oil to be swept out if the oil has a viscosity of 20 centipoise, and this is raised to only 858.1 days if the water viscosity is as high as that of the oil (taken here as 20 centipoise), the sand having a porosity of 0.2.

With regard to the problems of Section VI, we shall simply point out that the progressive distortions of the line of particles as they leave their initial positions are due to their entry into regions of ***non-uniform*** potential or pressure gradients. Thus, in the problem of the line drive into a well, the high gradients are localized about the perpendicular from the well to the initial position of the line. Since the fluid velocity is directly proportional to the pressure gradient, the fluid particles in these regions will move ahead of those in regions of lower gradients, thus giving rise to a "fingering" of the line towards the outlet well.

To interpret these traces of the line of fluid particles in a homogeneous system into their analogs in a real encroachment problem, it is sufficient, for qualitative purposes, to observe that these traces give a lower limit to the rate of encroachment and to the magnitude of the distortion (fingering) that would be observed in

[8] *A Theoretical Analysis of Water-flooding Networks*, M. Muskat and R. D. Wyckoff, Techn. Pub. No. 507, A.I.M.M.E. Pet. Div. (1933).

the non-homogeneous system. For, if the encroaching fluid is of lower viscosity (higher c) than that of the fluid displaced, the distortions will even be accentuated, since the replacement of the higher viscosity fluid will increase the gradients at the interface and hence the rates of distortion. This effect, however, will come into play only in regions of non-uniform gradients and hence only where the distortion, even in a single fluid system, is appreciable.[9]

[9] Although these considerations give a qualitative description of the nature of the effect of the viscosity difference between the encroaching and displaced liquids it must be remembered that in considering the quantitative effects on the simplified encroachment picture, one must distinguish carefully between the idealized representation in which it is supposed that the encroaching water completely displaces the oil originally in the sand, and the actual situation in which the oil is only partially replaced by the incoming water. In the former case, which is really based on the assumption of the complete miscibility of the two liquids, the porous medium will be the same after it is flooded by the water as before, the effect of the encroachment being simply the replacement of the oil by the water which will usually be of a lower viscosity. In truth, however, water and oil are by no means perfectly miscible and hence display a differential surface behavior which results in a retention by the sand of part of its original oil in spite of the flooding action of the encroaching water. This remnant oil will clearly lower the effective permeability of the flooded zone for the water which tends to pass through it, so that in the real encroachment problem the encroaching liquid is not only of a different viscosity than that of the liquid displaced, but moreover the flooded zone has a lower permeability than that still containing all of the original liquid. And although for analytical purposes this lowering of the original permeability due to the remnant oil may be immediately translated into an equivalent increase in viscosity of the encroaching water, it is important to note that this increase may well be so large as to more than counterbalance the fact that the water is strictly of lower viscosity than the oil. In fact it is quite possible that in actual encroachment systems the effective k/μ for the flooded zone is *less* than that in the oil zone in spite of the lower real viscosity of the water. Under these conditions, then, the idealized picture in which the viscosity difference was neglected will have to be modified to correspond to a flooding of the oil zone by a liquid of *higher* viscosity. The modifications outlined above for the case where the encroaching fluid is of lower viscosity will then be just reversed and fingering or cusping as well as the *rate* of encroachment will be retarded as compared to the results predicted for a system with no viscosity difference between the two fluids.

Finally, it is of interest to note briefly the relation of the general problem treated here and that usually occurring in practice in which the oil is saturated with gas and flows into the well by virtue of the expansion of the gas coming out of solution as well as by the "drive" of the encroaching water. Although the laws of flow of such "live" fluid systems have not yet been empirically established it is known that they possess pressure gradients which are much more concentrated about the outflow surfaces than those in the corresponding gas-free liquid systems. From this fact and the observation that at least part of the oil entering the well is due to direct displacement by the expanding gas, it is evident that the volume of water entry into a given region will be *less* than the volume of oil displaced and that there will, therefore, be a *retardation* in the encroachment with respect to that in a dead fluid system. Furthermore, the rate of water encroachment relative to the rate of oil depletion—the water entry per unit volume of oil produced—will be a minimum for high production rates which permit the evolution of the gas ("flush" production) and will increase and approach the value of 1 as the gas evolution is suppressed and the oil is displaced by the direct drive of the encroaching water. As to the shape of the water-oil interface, however, the same general features characteristic of the dead-fluid systems will also apply to the "live" fluid system, the more concentrated gradients in the latter simply giving the effect of a higher effective viscosity in the gas-oil zone than if the gas were in solution or were not present at all.

The writer is indebted to Mr. R. D. Wyckoff and Dr. A. E. Ruark for helpful discussions of this problem, to Dr. P. D. Foote for permission to publish the paper, and to Mr. M. W. Meres for the drawing of the curves and carrying through the calculations on which they are based.

An Experimental Study of the Motion of Particles in Systems of Complex Potential Distribution

R. D. WYCKOFF AND H. G. BOTSET, *Gulf Research and Development Corporation, Pittsburgh, Pa.*

(Received June 7, 1934)

A method is described whereby the motion of the interface between two liquids of equal viscosity and density moving through a porous medium in systems of complex pressure distribution may be obtained by the use of electrolytic models. Experimental results for certain simple cases are compared with results obtained analytically and shown to be accurate to within 95–98 percent. Illustrations are given of more complex cases showing the application of these models to some practical problems of oil field technology.

INTRODUCTION

THE flow of fluids through the porous sands or rocks which constitute the underground fluid reservoirs is of considerable practical interest since the production of water, oil or gas from wells tapping the sources depends upon such flow. It is, therefore, desirable that the detailed behavior of such fluids be known in order that the wells may be distributed and handled in the most advantageous manner. While it is clear that these underground porous strata are not homogeneous, and hence no exact analysis may be made in practice, yet a knowledge of the behavior of the fluids in a homogeneous porous system is an essential foundation upon which to base engineering technique.

Experiments show that at low velocities, porous flow is purely viscous with a gradual transition toward turbulence as the velocity exceeds a critical value for the particular medium. This is analogous to the *abrupt* transition from viscous to turbulent flow which obtains in open conduits when a critical velocity is reached. However, the permeabilities of the porous media encountered underground, are so low that any departure from purely viscous flow involves pressure gradients which are out of the range ordinarily encountered. Thus in these problems, it may be said that the flow of inelastic fluids obeys Darcy's law, the velocity being proportional to the pressure gradient; hence the solution of such problems involves the usual potential theory.

It is evident that the exploitation of underground fluid sources involves multiple well systems and in the case of certain "artificial" production methods commonly used in the oil industry, not only multiple output wells or sinks but also source wells into which are injected water or other fluids, the purpose of which is to augment the depleted natural drive whereby oil is forced into the output wells. Clearly the analysis of systems involving such complex arrays and shapes of sources and sinks may require considerable labor especially for the solution of certain phases of these problems. Since the results need not be of a high order of accuracy, the use of models is advantageous in many ways.

The phase of the flow problem with which this paper is concerned, involves the tracing of the history of a water-oil interface as the water encroaches from its source through the oil horizon and into the output wells. Clearly the analytical solution involves a complete knowledge of the potential distribution and streamlines. There are some cases of interest in which such a solution is not obtainable analytically, but in all cases the model provides reasonably accurate and interesting results.

DESCRIPTION OF THE MODEL

The model which is described here was developed to study the particular phase of porous flow problems mentioned above, that is, the tracing of a liquid front as it advances from its source to its outlet, as for example, the simple case of a liquid advancing from a line source into a point sink. For simplicity it will be tacitly as-

sumed that we are dealing with two liquids of equal density and viscosity, thus making the problem that of a homogeneous liquid.

Since for steady state viscous flow, the pressure distribution in a porous medium is exactly equivalent to the potential distribution in an electrical conducting medium, and since the velocity of a fluid particle is proportional to the pressure gradient just as the velocity of an ion is proportional to the potential gradient, a model utilizing ion migration should be exactly analogous to a homogeneous porous flow-system. It is clear, however, that such a model must fulfill two special requirements not ordinarily considered in electrolytic systems. Thus, since it is desired to trace the ion movement from its source to the output, that is, from the negative into the positive electrode, the model must be so arranged that only one type, say the OH-ion, is present in the system. It is also clear that the velocity of normal diffusion of the ions must be small compared with the velocity due to the applied potential difference. These requirements were the factors dictating the type of set-up described here. Because of the nature of the problem the model is somewhat simplified in that only two-dimensional systems are involved.

The requirement that the rate of normal diffusion of the ions be small requires that the electrolyte be held in a porous medium of low permeability. The use of gelatine would be ideal but preliminary experiments showed the method impossible because of hydration of the gel caused by chemical reaction and heating. Very little experimenting was done with the gel models for it was found that ordinary fine grained white blotting paper formed an excellent and convenient medium for the required purpose. Lack of time prohibited any search for a better material.

The blotting paper is soaked in a weak solution of potassium sulfate to which is added a generous quantity of a saturated solution of phenolphthalein. The blotter must be kept moist during the experiment but not too wet, for an excess of fluid causes rapid diffusion and destroys the sharp front caused by the potential gradient. It is also clear that the blotter model must be supported so that the saturation is uniform over its entire area, particularly any excess of the electrolyte along the edges must be avoided. The use of several glass point supports has been found satisfactory.

If two electrodes are placed in contact with the electrolyte and a potential difference is applied, hydroxyl ions move from the negative toward the positive electrode and their advance is indicated by the colored front due to the phenolphthalein. Simultaneously, however, hydrogen ions formed at the positive electrode move toward the negative electrode. If the H-ions are allowed to advance unhindered, they will meet the hydroxyl ions somewhere midway between the electrodes and progress of the indicator front will be stopped. Since it is essential that the hydroxyl ion front be followed right into the positive electrode or "sink," it is clear that either the "sink" must be located closer to the negative than to the positive electrode, or the hydrogen ions must be neutralized as soon as they are formed. The former method is extremely simple and may be used for certain network models. An example of this method is shown in Fig. 1.

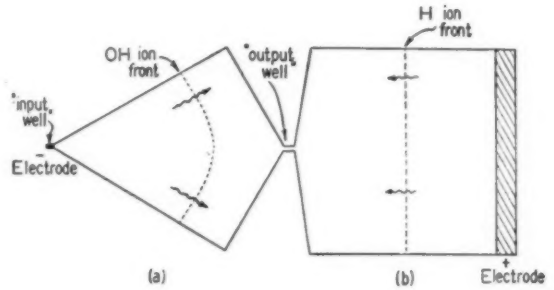

FIG. 1. Interference of the H ion with the OH ion between input and output is prevented by locating the output closer to the negative than to the positive electrode.

The portion of the model shown at (a) represents a sector of a "seven-spot" flooding network which will be described in detail later. The portion (b) is simply an extension of the electrolyte which has sufficient area to prevent the H-ions from reaching the output well before the colored OH-ion front has reached the output well. Some difficulty is experienced in operating this model due to the high current density and resultant heating at the constricted portion representing the output well. It is possible to add *small* quantities of electrolyte at this point during the process without seriously interfering with the

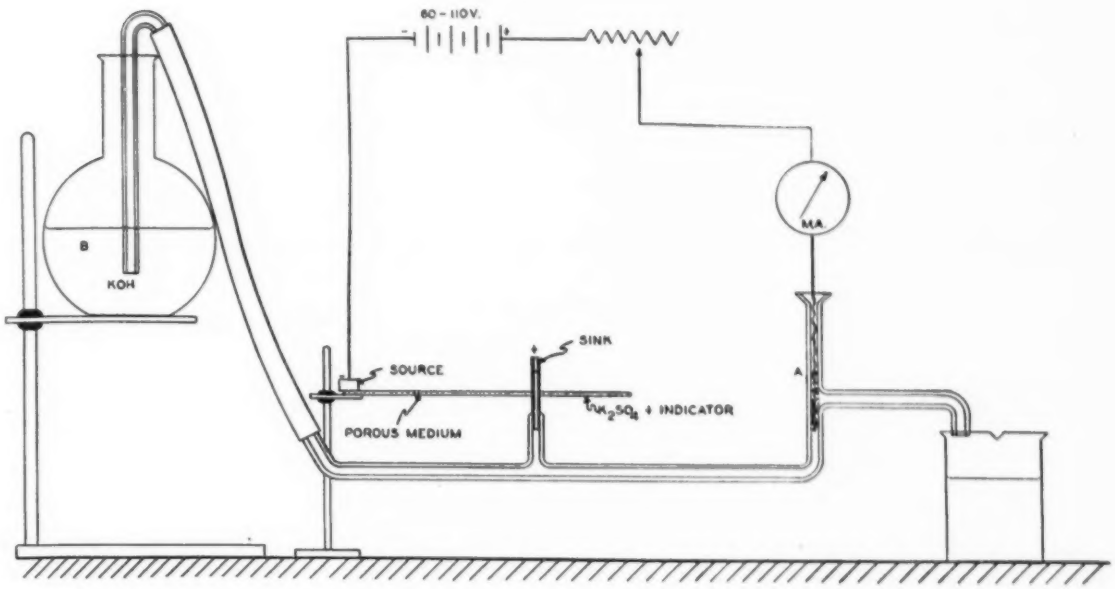

FIG. 2. Method for avoiding interference of H ion with the OH ion by neutralization at the anode.

advancing OH front since the direction of diffusion of the electrolyte is opposite to the advancing ions and the sharp definition of their advance is not disturbed.

The second method, involving neutralization of the H-ions is satisfactory in operation and is the method to be described in some detail. In this type of model the anode used is a porous tube about 1/16″ outside diameter fused into a glass tube as shown in Fig. 2. Suitable porous tubes are the small unglazed porcelain insulating tubes commonly used on thermocouple leads. The platinum electrode at *A* dips into an alkaline solution which is kept constantly flowing at a low rate from the reservoir flask *B*, into a beaker. Thus the porous tube (anode) is always alkaline and the hydrogen ions are neutralized immediately upon formation. The negative electrode consists of a metal strip or rod and may be of any shape required in representing the input boundary of the flow system. For the experiments described here it was either a point or a straight line. If the electrode is not a point, care must be taken that the contact along the entire edge is uniform. Slight variations in contact resistance along the edge of the electrode will cause serious distortions in the potential distribution and hence in the shape of the advancing front. The arrangement in Fig. 2 permits photographing various stages of the advance of the hydroxyl ions.

Other electrolytes than potassium sulphate may be used if desired but the same precautions regarding the position of the "sink" and the electrodes must be observed. The potassium sulphate-phenolphthalein combination used in these experiments was found to be decidedly superior to the other electrolytes tried in the preliminary experiments.

It is essential if the results are to be of value that the velocity of migration of the ions due to the potential gradient must exceed their normal diffusion velocity. The rate of diffusion in ordinary blotting paper is sufficiently low that this condition is satisfied when reasonably high potentials are applied. Potentials used were of the order of 20 volts per centimeter. When the distance between electrodes is about 4 cm and the potential gradients, the values mentioned above, the time required for the "flood" to advance from the source to the sink will be about thirty minutes, thus allowing ample time for taking as many photographs as desired.

PROOF OF ACCURACY OF MODEL

The study of the history of an advancing fluid front by purely analytical methods requires a

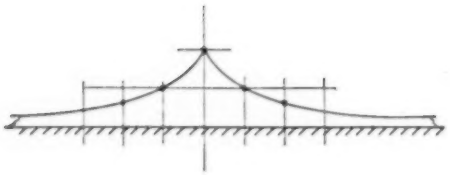

FIG. 3. Comparison of experimental and analytical results for the flow between an infinite line source and a single point sink.

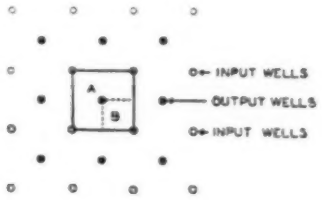

FIG. 5. "5-spot" array input and output wells commonly used in oil fields.

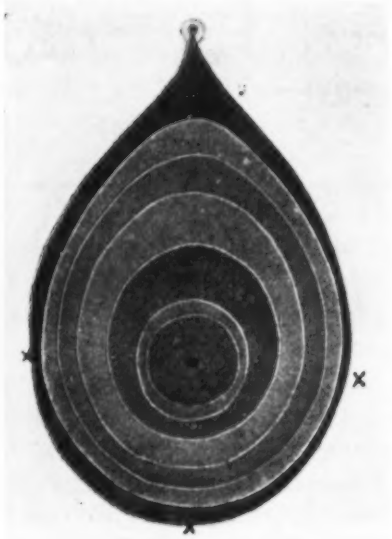

FIG. 4. Experimental results for the flow between a single source and sink in an infinite plane. Crosses are from the analytic solution.

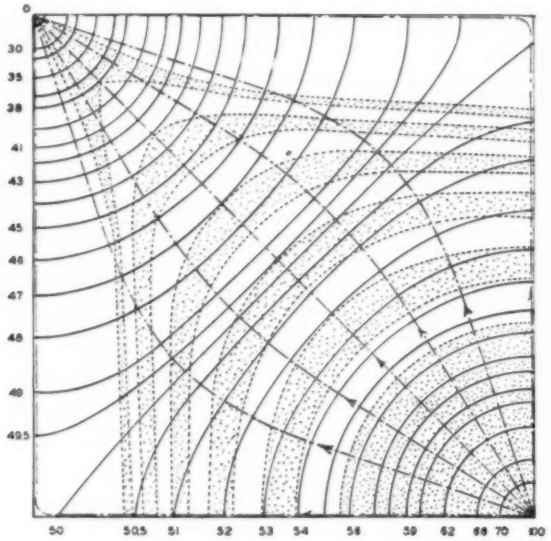

FIG. 6. Potential distribution over area *B* of Fig. 5 determined by metal conduction model.

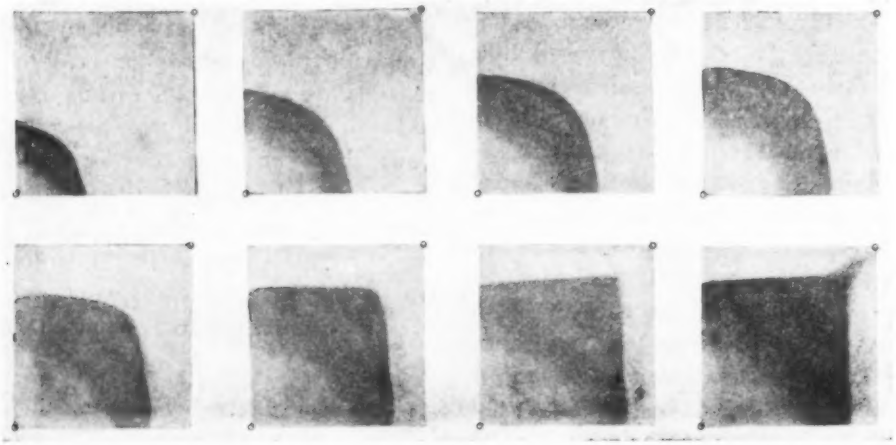

FIG. 7. Potential distribution over area *B* of Fig. 5 as determined by the electrolytic model.

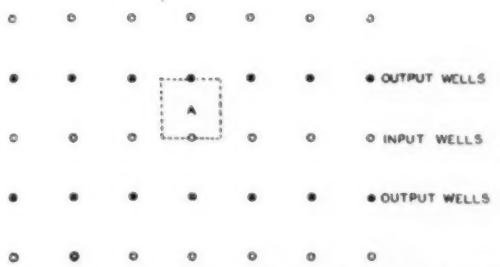

FIG. 8. "2-spot" flood arrangement of input and output wells used in oil fields.

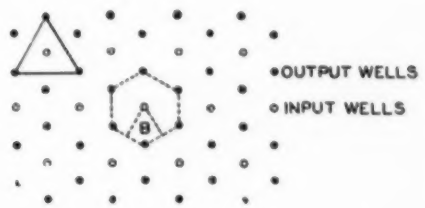

FIG. 10. "7-spot" arrangement of input and output wells.

complete solution involving both potential distribution and streamlines. Such analytical solutions had already been obtained for a few simple cases,[1] at the time this model was developed, and these solutions were used to determine the validity of the electrolytic model. The analytical results for an infinite line source and a single point sink in an infinite plane are shown in Figs. 11 and 12 (cf. 1). The various systems of curves in the figure show the potential distribution, the streamlines and various stages of the advancing front. The lines marked $t=0.05$, 0.10, etc., represent the positions of the advancing front at various stages of the flood. The line $t=0.33$ shows the position of the front the instant the flood reaches the output well.

Fig. 3 is obtained from the electrolytic model and is a tracing of the ion front the instant the hydroxyl ions reach the electrode or output well. The use of a large blotter gave conditions sufficiently close to an infinite plane that the distortion of the potential distribution was, for practical purposes, negligible. The points marked x are the corresponding points taken from the analytical solution and it is seen that the agreement is quite satisfactory.

Fig. 4 shows the experimental results for the flow between a single source and sink in an infinite plane. In this figure the photographs of various stages of the flood were cut out, superposed and then rephotographed. The crosses give the corresponding points of the analytical results (cf. 1, Fig. 14) showing fairly good agreement. The small discrepancies are due largely to the fact that the limited width of the blotter (necessitated by the mechanics of the experimental set-up) produced some distortion of the normal potential distribution with a resultant narrowing of the flood.

There was available a third check on the electrolytic model. This was a semi-analytical result obtained by actually determining experimentally the potential distribution between source and sink on a metal plate, from these, determining

[1] Morris Muskat, Physics **5**, 250 (1934).

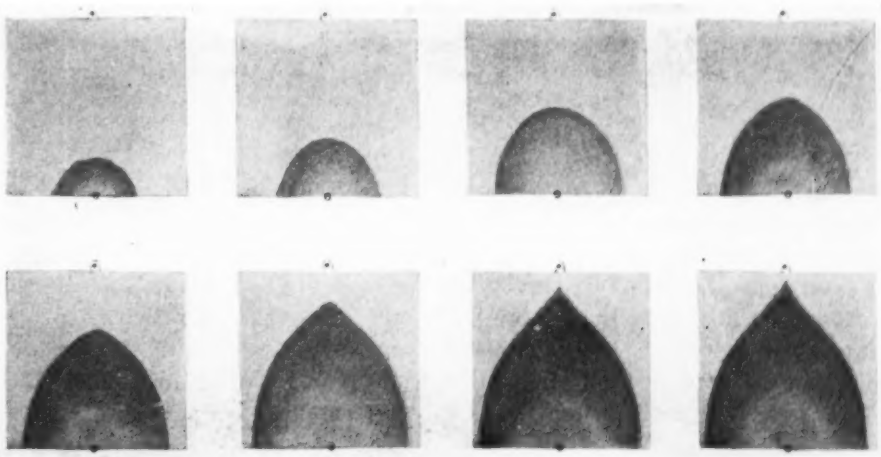

FIG. 9. Experimental results for the "2-spot" flood arrangement as obtained by the electrolytic model.

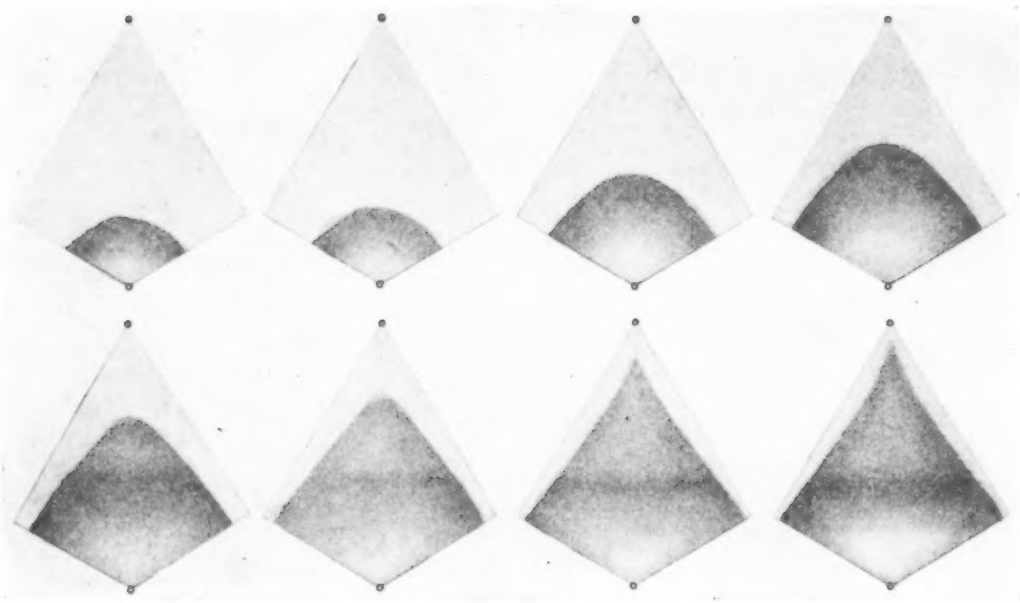

FIG. 11. Experimentally determined flow for the "7-spot" arrangement.

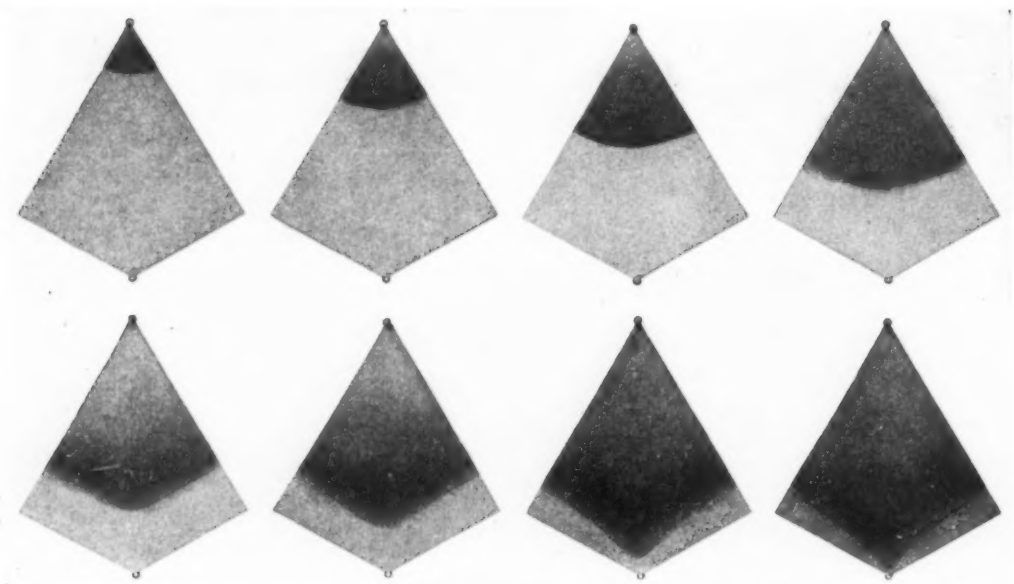

FIG. 12. Experimental results when input and output wells are reversed in the "7-spot" arrangement.

the streamlines, and thence by graphical methods the history of the advancing front. In oil fields in which water flooding is practiced, a more or less common distribution of input and output wells is what is known as the "5-spot" array. This is shown diagrammatically in Fig. 5. It will be seen immediately that with an extensive array of this type, a square (*A*, Fig. 5) at some distance from the edge of the area can be selected across whose boundaries there is no flow. In other words the boundaries are lines of symmetry. Furthermore a smaller portion of this square (*B*, Fig. 5) can be selected whose boundaries are also lines of symmetry. This smaller square contains at diagonally opposite corners one input and output well. Results from a study of the flood in this small

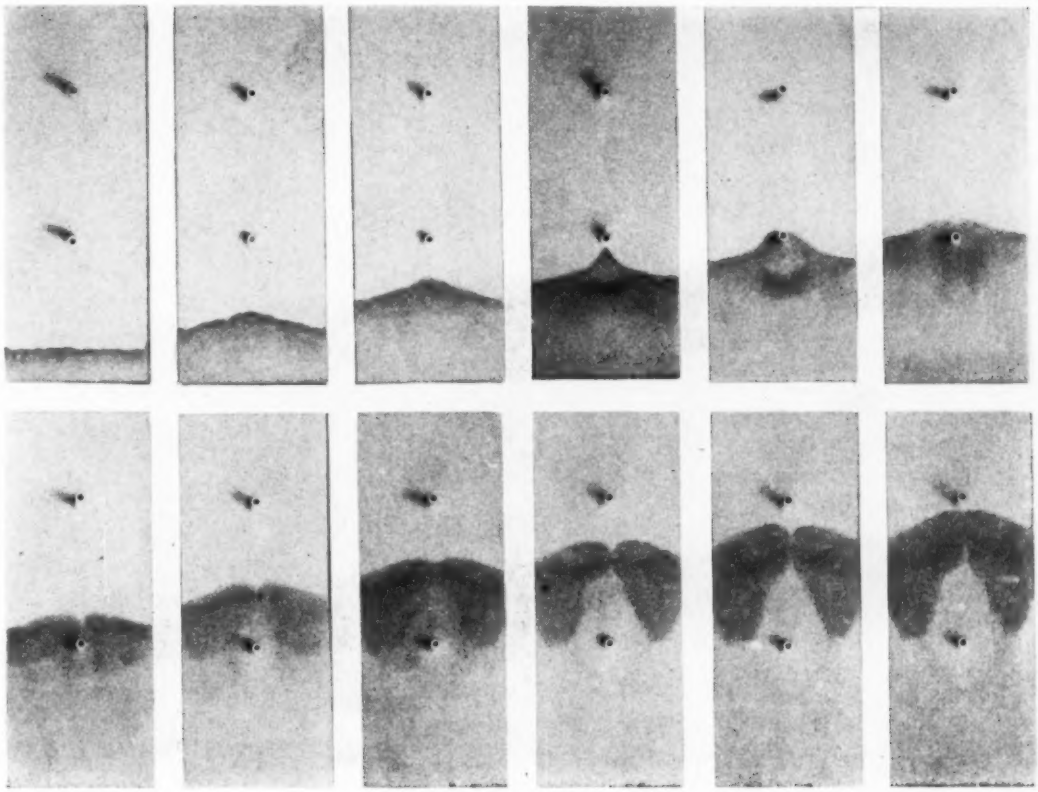

FIG. 13a.

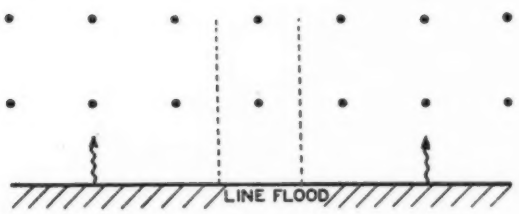

FIG. 13b.

FIG. 13. Line flood arrangement of input and output wells and experimental results for the flow.

square can then be applied to the entire field, except near the outer boundaries of the field where there will be some distortion due to unsymmetrical potential distribution at the limiting boundary of the array of wells.

The selection of a small section of a large network of input and output wells to be represented by an electrolytic model is not nearly as simple as it appears at first sight. Care must be taken that the boundaries of the model are lines of symmetry in the actual field case. If the boundary of the model is not a line of symmetry, distortions of the flood will be produced which will make the whole experiment meaningless. It would be possible for example, with a given electrode arrangement and spacing, to produce almost any desired shape of flood by varying the boundaries of the model. These shapes, of course, would be entirely useless for practical applications. The selection of lines of symmetry is sometimes very difficult and in many cases entirely impossible. Since a line of symmetry is really a limiting streamline, it is only in the very simple geometrical arrangements that these can be selected, as it

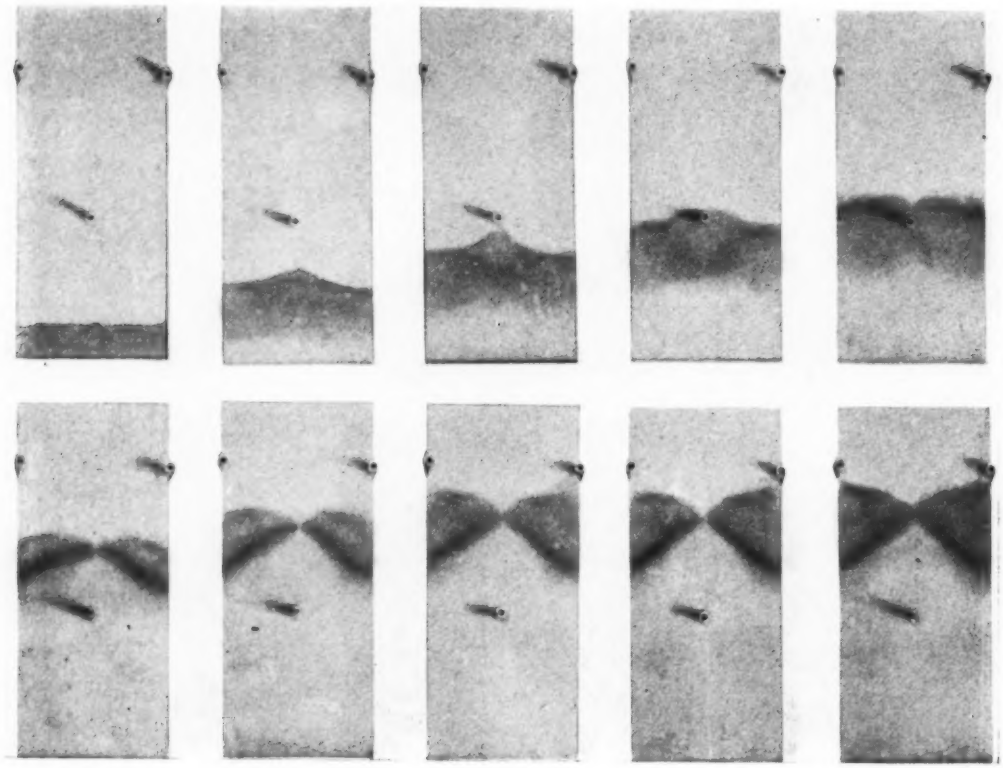

FIG. 14a.

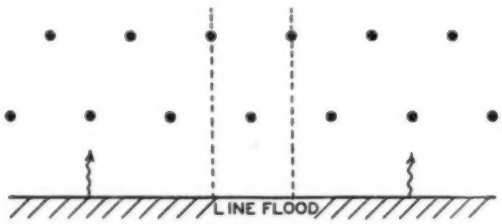

FIG. 14b.

FIG. 14. "Staggered" line flood arrangement of input and output wells and experimentally determined flow.

were, by inspection. The examples presented in this paper are all of this type. In the more complicated cases the selection of any stream line and hence any line of symmetry would involve the solution of the problem itself. In such cases where lines of symmetry cannot be selected with confidence, the only alternative would be to set up a very complicated system of input and output wells covering a considerable area around the region being studied and then to study only the central portion of this area.

The potential distribution over the area *B* (Fig. 5) was obtained by actual measurement in a metal conduction model; from this, by graphical method, the streamlines and the advancing front were obtained. These approximate results are shown in Fig. 6. Fig. 7 shows the results obtained on the electrolytic model. In comparing these photographs with the analytical results given above, the area flooded out at the instant the output well was reached was determined for each of the two methods, by a planimeter. In the analytical study[2] 72 percent of the total area was

[2] M. Muskat and R. D. Wyckoff, A.I.M.E. Technical Publication 507. *A Theoretical Analysis of Flooding Networks.*

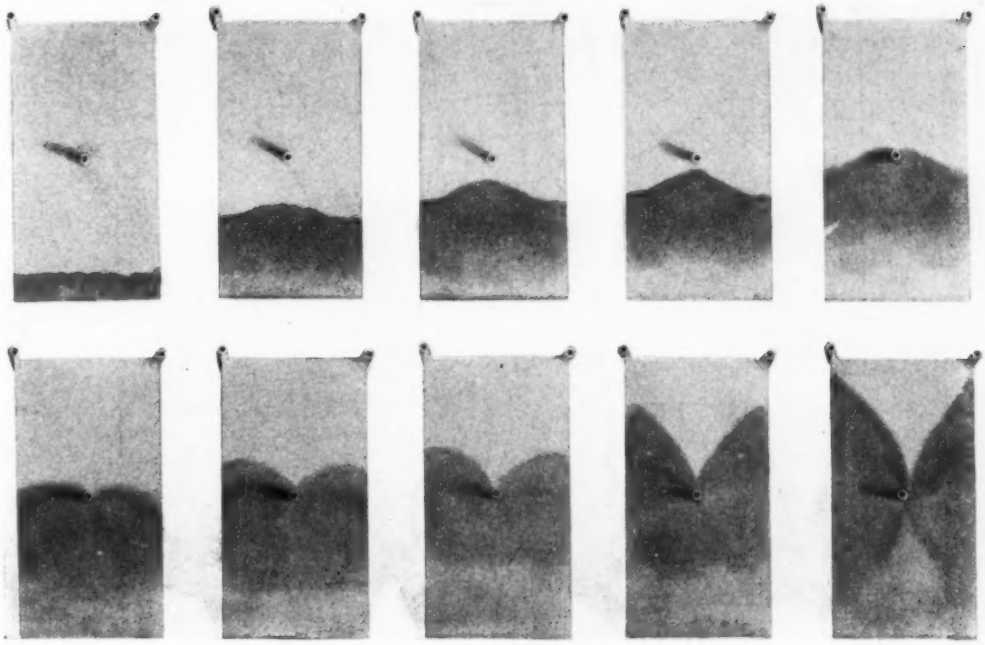

FIG. 15a.

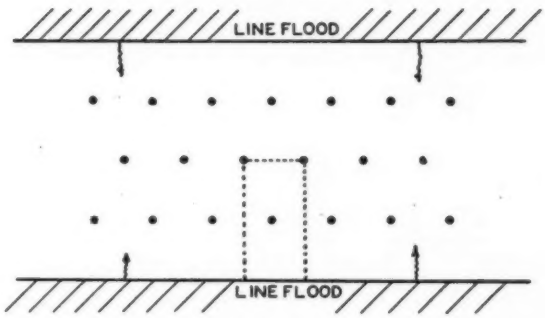

FIG. 15b.

FIG. 15. Double line drive into triple line of wells.

flooded out, while the result obtained from the electrolytic model was 73 percent. Thus the agreement of the two methods is again seen to be very satisfactory. Having thus shown that the electrolytic model is sufficiently accurate for practical purposes, we may now proceed to the consideration of more complex cases.

APPLICATIONS OF MODEL

In the "5-spot" flood discussed above the number of input equals the number of output wells. There is another distribution used in practice in which the input and output wells are equal in number. This is shown in Fig. 8 and, analogous to the 5-spot, may be called the "2-spot" flood. The boundaries of A (Fig. 8) are lines of symmetry, and the distance between lines of wells is equal to the distance between individual wells in a given line. Fig. 9 shows the experimental results for the 2-spot flood. The fractional area flooded out in this case is only 54 percent as compared with 57 percent obtained analytically.[2]

If the ratio of input to output wells is 2 to 1 we have what is called the 7-spot, since this ar-

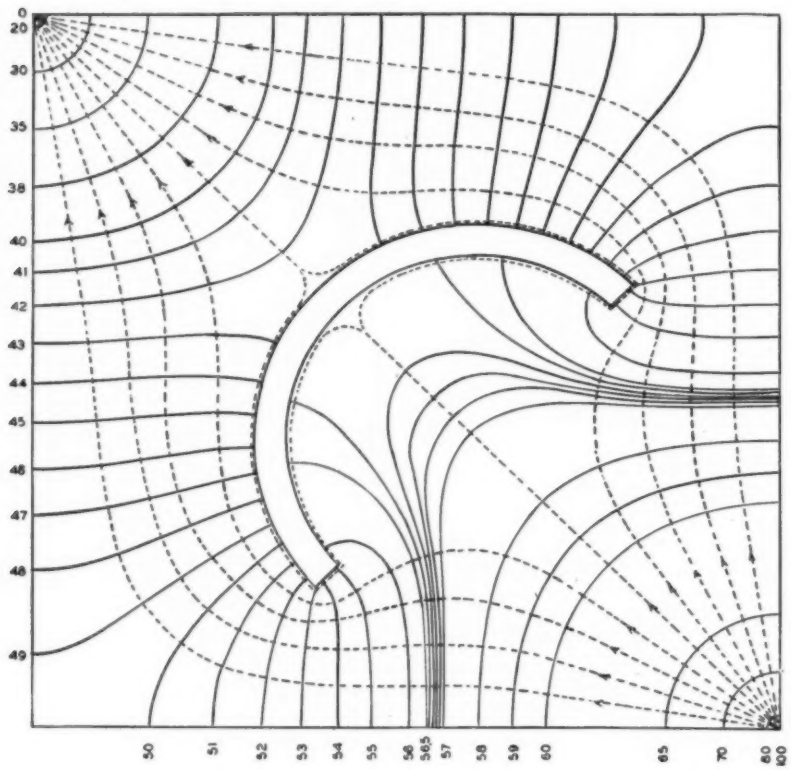

FIG. 16. Experimental solution of the same problem as in Fig. 6 except for the semi-circular obstruction.

rangement results in one input well at each corner of a hexagon (Fig. 10) with an output well at the center. The symmetry of this set-up allows us to select one input and one output well (*B*, Fig. 11) for the experiment, the results of which are shown in Fig. 11. The area flooded out in this case is 80 percent of the total. Now if the input and output wells were reversed so that the well at the center of the hexagon becomes the input and the others the output wells, there should be no change in the efficiency of the flood since the potential distribution is the same, but the shape of the flood should change. Fig. 12 shows how the experimental result verifies these conclusions. The flooded areas in these cases are 8 percent too high,[3] the discrepancy being due to the "flood" front approaching so close to the edges of the model that excessive diffusion occurs due to unavoidable concentration of the electrolyte at the edges.

[3] M. Muskat and R. D. Wyckoff, reference 1. (It should be mentioned that at the time this work with models was done, the analytical results for this and following cases were not available.)

One important phenomenon connected with oil production is that of the encroachment of edge-water into the oil fields. As pressure gradients are established across the field due to the oil production, the water along the edge, being under a maintained hydrostatic pressure, gradually advances into the field replacing the oil removed. A knowledge of the progressive changes in the shape of this advancing oil-water interface is of great importance. For purposes of this study the original oil-water interface has been assumed to be a straight line. If the assumption is made that there are two rows of producing wells parallel to the original water front, the arrangement shown in Fig. 13 is obtained. Lines of symmetry may here be easily selected and the advance of the flood over one section is shown in Fig. 13. In this experiment the potential was maintained on both wells during the entire experiment. In practice it is probable that flow into the first well would be stopped shortly after the water reached the well. If this were done, the advancing front would show a sharper cusp into the second well, since

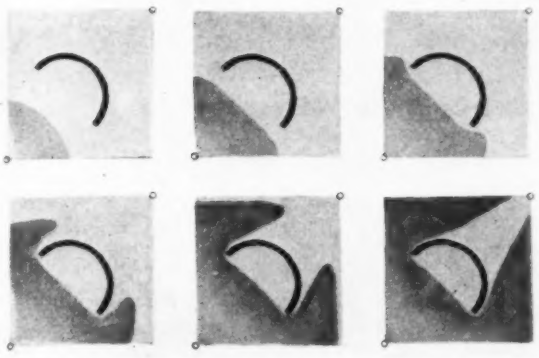

FIG. 17. Same problem as Fig. 16 solved by the electrolytic method.

the potential distribution would be changed when the first well was shut off.

If the wells of the second line, instead of being placed directly behind those of the first are placed on the perpendiculars bisecting the distance between the first line wells, the arrangement of Fig. 14 is obtained. The effect of thus "staggering" the second line of wells may easily be seen by comparing Figs. 13 and 14. It would appear that the total area flooded out is approximately the same for either arrangement. In this case the lines of symmetry are two parallel lines extending an indefinite distance from the line flood source, which means that the blotter used in the experiment must extend a considerable distance beyond the second row of electrodes. In the illustration it has been cut off at the wells merely to conserve space in the figure. If the experiment were made with a blotter cut off at the second row of wells, the altered potential distribution would produce the result shown in Fig. 15 and would represent the field problem indicated in the figure. Here the cusping into the second row of wells is very sharp and the area flooded out considerably less than in the preceding figure. This illustrates the importance of accurately selecting the boundaries of an area to be used in a given experiment.

As an example of some of the more complex problems whose solution is easily and directly obtainable by the electrolytic model, Fig. 16 and 17 may be cited. This is exactly the same set-up as was used for the 5-spot flood of Fig. 7, except than an obstruction (consisting merely of an opening cut in the blotter) was placed between the source and sink.

The effect of the obstruction on the potential distribution and streamlines and the shape of the flood (features impossible of calculation) are readily seen in Figs. 16 and 17, respectively. From this and the other illustrations given it is clear that where an accuracy of about 95–98 percent is sufficient, the electrolytic model as well as current conduction models are useful in obtaining solutions to certain problems which are very laborious or even insoluble by the usual analytical methods of potential theory.

The Radium Content of Some Connate Waters

HOLBROOK G. BOTSET, *Gulf Research and Development Corporation, Pittsburgh, Pa.*

(Received June 6, 1934)

A method is described for measuring the radium content of natural waters. The method is shown to be accurate to within a few percent for concentrations as low as 100×10^{-12} g Ra per liter. Results are given of measurements on several connate waters from mid-continent oil fields. A correlation between radium and calcium content of the waters is indicated. Samples taken from the same wells over a period of several months indicate that the radium concentration is a definite property of a connate water.

IN the course of some measurements of the radon content of soil gases[1] it became apparent that there was a correlation between this and the heavy mineral content of the soil.[2] The consideration of these results immediately suggested that there might be some significance to the radium content of connate waters and that it should be considerably greater than in surface waters. Presumably connate waters of the deeply buried rocks are at present practically at rest, hence their salt content is more or less characteristic of the geological formation in which they occur. In fact this method is quite generally used in the oil industry in correlating sources of various oil field waters. Since connate waters frequently have several times the salt content of sea water, radium should be present in sufficient quantity to be fairly readily measured.

A survey of the earlier literature on the radioactivity of natural waters[3] does not throw much light upon the question, since in practically all of these measurements, no distinction was made between radon and radium content of the waters. Most of the measurements were of the radon content of the thermal springs and no attention was paid to the radium. Although many of these earlier results are given in Emans per liter or Mache units it has become generally customary to express the radioactivity of waters in grams of radium $\times10^{-12}$ per liter. For example sea water from the Gulf of Mexico was found to contain 1.7×10^{-12} g of radium per liter. One of the highest concentrations listed in the *International Critical Tables* is 100×10^{-12} g Ra per liter for the water of Fairmont Springs in British Columbia. As far as could be determined, the literature contained no data on the radioactivity of connate waters.

Since our interest was in the radium content of the water the first problem was to determine whether or not the amount of radium normally present could be detected and measured with some degree of accuracy by any fairly simple laboratory method. It was felt that the method should be sensitive enough to measure the radium in not more than two gallons of water, as a larger sample than that would generally be inconvenient to handle.

It was decided to try the method of precipitating the radium with barium as a radium-barium sulfate, filtering, fusing with $NaKCO_3$ and dissolving in HCl. The radon produced in the sealed solution is then determined by the Lind method.[4] This method is commonly used in measuring appreciable quantities of radium, however, the extremely small concentrations found in natural waters would lead one to question the efficacy of this method of removing the radium from the water sample. Ebler and

[1] H. G. Botset and P. Weaver, Physics **2**, 376 (1932).
[2] R. W. Clark and H. G. Botset, Bull. A.A.P.G. **16**, 1349 (1932).
[3] A. Gockel, Die Radioaktivität von Boden und Quellen, Vieweg, Braunschweig. J. Joly, Radioactivity and Geology, Constable, London, 1909; International Critical Tables, Vol. 1; *Radioactivity of Natural Waters*, Scholl, Univ. of Illinois Bull. Vol. 15, No. 11, Water Survey Series 14, p. 114.
[4] Analytical method for certain metals, Moore, Lind, etc., U. S. Bureau of Mines, Bull. 212.

Van Rhyn[5] and others[6] have shown that if barium be added to a solution containing minute quantities of radium and be precipitated quantitatively as sulfate the radium is also carried down practically quantitatively, regardless of whether or not the solubility product of the radium sulfate is satisfied.

To precipitate the radium in these experiments 10 cc of a solution containing 2.5 g of $BaCl_2 \cdot 2H_2O$ are added to the water sample. To this, 15 cc of $1:1 H_2SO_4$ are added, thus precipitating the barium quantitatively and leaving an excess of sulfate ions in the solution. The barium sulfate thus precipitated, settles readily, allowing most of the water to be decanted or syphoned off and the remainder is filtered. The precipitate is then dried, and fused with an excess of sodium and potassium carbonates in a small nickel crucible. The fused mass is leached with water until soluble sulfates and carbonates are dissolved, and filtered. The insoluble barium carbonate carrying the radium, after washing with sodium carbonate solution to freedom from sulfates, is dissolved in HCl. The resulting solution is sealed in an Erlenmeyer flask and the radon allowed to accumulate for a few days. The radon is measured by the Lind method and calculated to full equilibrium value.

Test of Sensitivity of Method

In order to determine the sensitivity of the method, samples were obtained from a few local streams and springs. A sample of water from the laboratory tap (source—Allegheny River) showed a radium content of 143×10^{-12} g per liter. Samples from two springs contained 66 and 15×10^{-12} g Ra per liter. A sample from another spring contained 31×10^{-12} g per liter.[7] These results indicated that the method was sufficiently sensitive for the purpose at hand. There remained the question of its accuracy: How much of the radium present was being measured? In order to answer this question, a two gallon sample of distilled water to which had been added, from a standard solution, 355×10^{-12} g of radium gave, by measurement, 334×10^{-12} g or 94 percent of the actual radium present, another solution containing a pitchblende sample of 2560×10^{-12} g of radium gave, upon measurement, 2520×10^{-12} g or 98 percent of the actual radium present. All of the reagents used in the precipitation and separation of the radium from solution were carefully tested for radioactivity and a suitable correction applied to the radium measurements.

It was thought, from these results that the method of measurement was both sufficiently sensitive and sufficiently accurate for the study of connate waters.

Method of Obtaining Field Samples of Connate Waters

Two gallon samples were collected at well heads or from settling tanks. The barium sulfate precipitation was made in the field with tested chemicals sent from the laboratory. The precipitate was allowed to settle and most of the supernatant liquid syphoned off. The resulting sample, about a liter in volume, was then shipped to the laboratory for measurement. In practically every case duplicate samples were taken.

The connate water samples were all taken from producing oil wells in Oklahoma and consequently represent types of water found in productive oil sands.

Experimental Results

The first group of samples was obtained from three different pools in Oklahoma. Table I gives the results of the measurements together with the chemical composition of the waters.

It may be seen from this table that the radioactivity of these waters is very high compared to that of surface waters. A large variation will be noted between the radium contents of the two samples from the Siliceous Lime. The thickness of the siliceous lime beds and the location of the oil producing zones within them are extremely variable.

In order to determine the constancy of the radium content of connate waters, samples were taken from these same wells four months later and their radium content determined again. The results of these measurements are shown in

[5] Ebler and Van Rhyn, Ber. d. deutsch. Chem. Gesell. **54**, 2696 (1921).

[6] F. E. German, J. Am. Chem. Soc. **45**, 1615 (1921); Levin, Phys. Zeits. **10**, 322 (1909).

[7] Comp. St. Meyer and Schweidler, Radioaktivität, Table p. 570.

TABLE I. *Radioactive content and chemical composition of waters from three pools in Oklahoma.*

Sample No.	1	2	3	4	5
Pool	Oakhurst	Boston	Boston	Glenn Pool	Glenn Pool
Depth	2156′	2294′	2649′	1420–1527′	2362′
Formation	Siliceous Lime	Bartlesville	Siliceous Lime	Glenn Sand	Wilcox Sand
g of* Ra$\times 10^{12}$	828	570	1937	475	540
Na (mg per liter)	57442	52829	56203	42251	50294
Ca (mg per liter)	11200	10820	13900	5680	8140
Mg (mg per liter)	2368	1872	2172	2548	1774
SO_4 (mg per liter)	—	—	82	60	63
Cl (mg per liter)	115200	106000	117500	82250	97000
HCO_3 (mg per liter)	67	48	61	520	87
Total solids	186277	171569	189918	133309	157298

* Although only one value is given for each well it is the average of two samples, the duplicate samples checking from within 1 to 4 percent.

TABLE II. *Constancy of the radium content of connate waters.*

Sample No.	1	2	3	4	5
Depth	2156′	2294′	2649′	1420–1527′	2362′
g Ra$\times 10^{12}$ (October)	836	581	1956	471	539
	820	560	1908	480	542
g Ra$\times 10^{12}$ (February)	814	549	2375	513	564
				510	
Ave. of two October samples	828	570	1937	475	540
% difference between ave. of October samples and Feb. samples	−1.7	−3.8	+22.0	+7.7	+4.3
Ave. of all three samples	823	563	2079	488	549
% maximum deviation from average of three samples	1.55	3.1	14.2	5.1	2.7

Table II, in which they are also compared with the results of the earlier measurements.

It will be seen from this table that with the exception of sample No. 3, the radium content of the waters examined varied but little during the four month interval, which indicates that the radium content of a water may be considered as a constant, characteristic of the water, for a period of at least several months. The variation in sample No. 3 may be due to some small casing leak in the well which permitted the entry of contaminating water from shallow horizons at the time the earlier samples were taken.

Several samples of water were obtained from two oil fields in the Mid-continent area, Glenn Pool and Burbank. The radium content of these waters as well as their Ra/Ca and Ra/Mg ratios are shown in Tables III and IV.

Table V gives a summary of results on some waters from the general Cromwell and Seminole areas, and also includes for purposes of comparison, averages of the results given in Tables III and IV. In averaging Table IV only the results of the last four samples are used as these seem to be the only ones which may be regarded as really characteristic of the formation from which they were taken and even these are open to some question as will be mentioned later.

TABLE III. *Radium content of waters of Glenn Pool.*

Sample No.	Ra($g/L\times 10^{12}$)	(Ra/Ca)$\times 10^9$	(Ra/Mg)$\times 10^9$
1	505	0.103	0.2135
2	531	0.111	0.202
3	519	0.0766	0.249
4	527	0.1063	0.208
5	558	0.111	0.212
6	447	0.158	0.343
7	499	0.101	0.191
8	499	0.077	0.224
Ave.	511	0.106	0.230

TABLE IV. *Radium content of waters of Burbank.*

Sample No.	Ra($g/L\times 10^{12}$)	(Ra/Ca)$\times 10^9$	(Ra/Mg)$\times 10^9$
1	301	0.060	0.342
2	1650	0.170	0.755
3	708	0.0444	0.302
4	911	0.057	0.409
5	127	0.0385	0.229
6	1588	0.132	0.602
7	2004	0.134	0.611
8	1448	0.133	0.686
9	1913	0.132	0.531
Ave.	1730	0.133	0.608

TABLE V. *Summary of results on waters from the general Cromwell and Seminole areas.*

Sand	Cromwell	Upper Cromwell	Wilcox	Second Wilcox	Simpson	Hunton Lime	Glenn	Burbank
Ra($g/L\times10^{12}$)	470	1286	293	320	208	552	511	1730
Ra/Ca($\times10^{9}$)	0.0569	0.1656	0.0304	0.0347	0.0240	0.0710	0.106	0.133
Ra/Mg($\times10^{9}$)	0.331	0.646	0.138	0.187	0.098	0.233	0.230	0.608
No. of Samples from each sand	4	1	6	2	1	1	8	4

It may be seen from these results that the radium content of a connate water is as characteristic of that water as are the other chemical elements present.

There is one limiting factor which will control, to a certain extent, the radium concentration, and that is the amount of sulfate present in the water. The solubility product constant for radium sulfate is so extremely low that only a very small amount of sulfate is sufficient to precipitate some radium from solution. Consequently only sulfate free waters would be expected to contain large amounts of radium. One of the characteristics of connate as distinguished from surface waters is the small concentration or absence of sulfate ion, particularly in waters associated with petroleum fields.

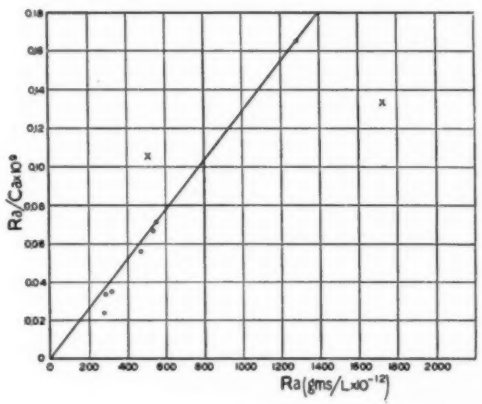

FIG. 1. Radium contents of the waters of Table V plotted against their Ra/Ca ratios.

In Fig. 1 are plotted the radium contents of the waters of Table V against their Ra/Ca ratios. There appears, with two exceptions, to be a linear relation between these quantities. The two exceptions are the Glenn and Burbank waters. The water from the Glenn oil field may show an erratic ratio due to the high content of carbonate and also the presence of sulfate. Both of these ions would tend to modify the calcium and radium content. In the case of the Burbank water, practically all the samples which could be used were taken from test tanks. The Ra/Ca ratio is low. Although the radium content of these samples appears high, it may still be less than the real radium content of the water, due to precipitation in the test tanks from which the samples were taken.

CONCLUSIONS

The radium content of a connate water appears to be as definite a characteristic of that water as is any of the other salt constituents. The determination of the radium content of a water might be of use in correlating or differentiating waters in cases where the ordinary salt content does not permit such correlation or differentiation

If connate waters, as is generally believed, are derived from ancient seas, the enrichment in radium (uranium)[8] content, has in general been far greater than the enrichment in other solids. This observation, together with the fact that the atomic weight of uranium is much greater than that of the ordinary elements present in the waters, might be of use in considering certain theories proposed regarding diffusion of salts through waters during geologic time. To study this question a great number of water samples from various sands and areas would have to be measured.

If helium is of radioactive[9] origin the study of

[8] V. Chlopin and W. Vernadsky, Zeits. f. Electrochemie **38**, 527 (1932), found some Russian waters in which the uranium was not in equilibrium with the radium content. Some other waters, however, did have the equilibrium concentration of uranium. The excess of radium in the Russian waters was attributed to the ability of waters to dissolve appreciable quantities of radium from rocks at elevated temperatures.

[9] *Helium Content of Natural Gas*, G. S. Rogers, U.S.G.S. Professional Paper 121.

the radium content of waters from formations in helium producing areas should throw considerable light on this subject.

The author realizes clearly that the number of samples discussed in this paper (about fifty) is much too small to permit drawing any very definite conclusions, yet it is felt that these results do indicate certain possibilities in the use of the radium content in correlating and studying connate waters and it is with the hope of suggesting and stimulating further studies along this line that the above material has been presented.

The author wishes to acknowledge gratefully the assistance of Mr. L. C. Case of the Gypsy Oil Company who obtained the water samples and made the chemical analyses, and also to express appreciation for the helpful guidance of Dr. A. E. Ruark under whose direction most of the work was carried out. Thanks are due to Dr. Paul D. Foote for permission to publish this paper.